

AN EXPLORATORY STUDY OF MUSIC TEACHER EVALUATION PRACTICES  
IN MULTIPLE STATES WITH RACE TO THE TOP FUNDING:  
K–12 MUSIC EDUCATORS' EXPERIENCES, PERSPECTIVES,  
AND RECOMMENDATIONS

by

LISA DIANE MARTIN

B.M.E., University of Illinois, 2005

M.M.E., University of Colorado, 2010

A dissertation submitted to the  
Faculty of the Graduate School of the University of Colorado  
in partial fulfillment of the requirement for the degree of  
Doctor of Philosophy  
Department of Music Education

2014

This dissertation entitled:  
An Exploratory Study of Music Teacher Evaluation Practices  
in Multiple States with Race to the Top Funding:  
K-12 Music Educators' Experiences,  
Perspectives, and Recommendations

written by Lisa D. Martin  
has been approved for the Department of Music Education

---

James R. Austin, Ph.D., Committee Chair

---

Margaret H. Berg, Ph.D.

---

James M. Dunn

---

David A. Rickels, DMA

---

Jennie Whitcomb, Ph.D.

Date \_\_\_\_\_

The final copy of this dissertation has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB Protocol # 14-0029

Martin, Lisa D. (Ph.D., Music Education)  
An Exploratory Study of Music Teacher Evaluation Practices in Multiple States with  
Race to the Top Funding: K–12 Music Educators’ Experiences, Perspectives, and  
Recommendations  
Dissertation directed by Dr. James R. Austin

### **Abstract**

The purpose of this study was to examine K-12 music educators’ perspectives of current teacher evaluation practices designed to comply with Race to the Top initiatives implemented in four states. A total of 288 music teachers from Florida, North Carolina, Rhode Island, and Tennessee completed a 23-item online questionnaire in the spring of 2014. Participants responded to closed-ended items regarding their most recent professional evaluation, including (a) measures of performance considered in the evaluation process, (b) stakeholders involved in evaluative decisions, and (c) outcomes associated with evaluation results. Additional research questions focused upon K–12 music educators’ perceptions of the fairness, clarity, and utility of evaluation processes and how teacher evaluation impacted daily classroom practice.

Closed-ended items were subject to descriptive analyses, and paired samples *t*-tests revealed relationships between music educators’ lived and idealized notions of teacher evaluation requirements. Group differences in teacher perceptions as related to gender, area (e.g., general, instrumental, choral), and level (e.g., K–5, 6–8, 9–12) were examined using multivariate analyses of variance (MANOVA). Responses to two open-ended items, focusing upon evaluation’s impact upon music teaching and learning, as well as evaluation’s influence on music education as a profession, were analyzed, coded, and organized into patterns.

Descriptive findings suggested that participants favored evaluation systems emphasizing professional development outcomes (e.g., to guide improvement of teachers’ skills) over those emphasizing professional status outcomes (e.g., to discharge incompetent teachers). Overall,

participants had slightly negative perceptions of teacher evaluations, sensing little direct benefit to teaching and learning. Music educators also lacked clarity regarding evaluation requirements and outcomes. Although participants supported the consideration of multiple measures of teacher performance in the evaluative process, they most valued classroom observations. Group difference analyses suggested that female elementary general music teachers responded most favorably to teacher evaluation.

Open-ended item analyses revealed participants' distinct concerns about time demands, lack of evaluator expertise in music, and overall applicability of the evaluation process to music teaching and learning. Many participants also believed the evaluation process diminished the professional status of music education. Implications for teacher evaluation policy and recommendations for research are discussed.

## **Acknowledgments**

Foremost, I extend my sincerest gratitude toward my advisor, Dr. James Austin, whose support made this document possible. Your guidance has impacted me immeasurably, and your humor has sustained me.

To Dr. Margaret Berg, thank you for contributing to my development as a teacher educator and scholar. You are an exceptional role model.

I would also like to express appreciation toward my other committee members: to Michael Dunn, for your constant encouragement and friendship, to Dr. David Rickels, for your dependable counsel and creative insights, and to Dr. Jennie Whitcomb, for asking questions that challenge my understandings and inspire my curiosity. I thank each of you for your feedback and support throughout this project.

To all of my teachers, mentors, friends, and colleagues throughout the years – every day, I appreciate the gifts you have shared with me.

Finally, to my parents, John and Susan Martin, thank you for unquestionably supporting me throughout this journey. I am beyond fortunate to have you in my life.

## Contents

### Chapter

<b>I. Introduction</b> .....	1
The Accountability Movement.....	1
History of testing and accountability in the United States.....	2
Race to the Top.....	5
Section summary.....	7
Teacher Evaluation: Conflicts and Concerns.....	8
Research Problem.....	9
Teacher Evaluation: Theories of Action and Motivation .....	12
The Status of Music Education in the United States.....	16
The challenge of evaluating music teachers.....	18
Evaluative instrument.....	18
Evaluator expertise.....	20
Music Education’s Response to Evaluation Initiatives.....	21
State-Level Approaches to Evaluation.....	23
Tennessee.....	24
North Carolina.....	25
Rhode Island.....	27
Florida.....	29
Need for the Study.....	33
Study Purpose and Research Questions.....	34

Definitions.....	35
Delimitations.....	36
<b>II. Review of Related Literature.....</b>	<b>38</b>
Introduction.....	38
Perceptions of Teacher Evaluation in General Education.....	40
Teacher evaluation functions.....	41
Measures of teacher performance.....	44
Student growth measures.....	47
Evaluator credibility and competence.....	48
Fairness, clarity, and utility.....	50
Educators' perceptions of RttT evaluation policy.....	56
Section summary.....	58
Evaluation and Professionalism.....	58
Section summary.....	60
Music Educators' Perceptions of Teacher Evaluation.....	61
Evaluative instrument and procedures.....	61
Evaluator training and credibility.....	65
Section summary.....	69
<b>III. Methodology.....</b>	<b>71</b>
Introduction.....	71
Sampling.....	71
Data management company option.....	74
NAfME national office option.....	75

State MEA option.....	75
Rationale for Survey Methodology and Teacher Self-Report Data.....	77
Online surveys.....	78
Survey platform.....	81
Teacher self-report data.....	81
Survey Responses.....	82
Questionnaire Item Pool Development.....	85
Preliminary study preparation: Commitment and morale.....	85
Closed-ended items describing the evaluation process.....	86
Closed-ended items on teacher perceptions.....	90
Open-ended items.....	92
Open-ended item development.....	93
Demographic data.....	96
Procedures.....	96
Pilot testing and pre-data collection.....	96
Data collection.....	97
Data Analysis.....	98
<b>IV. Results.....</b>	<b>99</b>
Introduction.....	99
Section A: Closed-Ended Items.....	100
Participant Demographics.....	100
Characteristics of the Evaluation Process.....	103
Evaluative purpose.....	103

Measures of teacher performance.....	107
Evaluative personnel.....	111
Music Teachers' Perceptions of Teacher Evaluation Qualities, Benefits, and Impact.....	116
Evaluation qualities.....	116
Evaluation benefits.....	118
Evaluation impact.....	119
Relationships Among Teacher Perception Subscales and Teaching Experience.....	121
Analyses of the Effects of Individual Differences on Teaching Perceptions.....	122
Section B: Open-Ended Items.....	126
Open-Ended Data Collection and Preparation for Analysis.....	126
Coding Open-Ended Responses.....	127
Verification strategies.....	127
Open-Ended Item #1.....	130
Theme: Impact.....	130
No impact.....	132
Negative impact.....	132
Positive impact.....	133
Theme: Concerns.....	134
Time concerns.....	137
Evaluator concerns.....	137
Fit concerns.....	138

Theme: Teacher reaction.....	139
Jumping through hoops.....	141
Self reliance.....	141
Desire for more appropriate tool.....	142
Section summary.....	142
Open-Ended Item #2.....	142
Responses and major themes.....	143
Theme: Impact on professional status.....	143
No impact.....	146
Negative impact.....	146
Diminished status of music educators.....	146
Diminished status of all educators.....	147
Negative impact upon teaching.....	147
Morale.....	148
Positive impact.....	148
Elevated status of music educators.....	148
Improved teaching.....	149
Theme: Concerns and professional status.....	149
Theme: Reactions and professional status.....	152
Leaving the profession.....	154
Improved process.....	154
Section summary.....	154

<b>V. Summary and Conclusions</b> .....	156
Summary.....	156
Discussion.....	160
Elements characterizing teacher evaluation.....	160
Evaluation purpose.....	160
Evaluation purpose: what <i>is</i> versus what <i>should be</i> .....	162
Measures of teacher performance.....	163
Measures of teacher performance: what <i>is</i> versus what <i>should be</i> .....	165
Evaluator role.....	166
Section summary.....	167
Perceptions of the music teacher evaluation process.....	168
Evaluation qualities.....	168
Evaluator characteristics.....	170
Section summary.....	172
Perceived benefits and impact of music teacher evaluation.....	173
Benefits of teacher evaluation: Group differences.....	176
Section summary.....	177
Music education as a profession.....	178
Implications and Recommendations.....	180
Suggestions for policymakers.....	180
Teacher evaluation design emphasis.....	180
Localizing control.....	181

Suggestions for music education leadership.....	181
Mutual accountability.....	181
Engaging stakeholders.....	182
Measuring music-specific student growth.....	183
Observations by music experts.....	185
Suggestions for classroom music educators.....	186
Study Limitations and Recommendations for Future Research.....	187
Conclusion.....	189
References.....	190

**Appendices**

A. Invitation for State MEA Participants.....	211
B. Introductory Email Script.....	212
C. Reminder Email Script.....	213
D. Music Teacher Evaluation Inventory.....	214
E. Institutional Review Board Approval.....	223
F. Pilot Version of Music Teacher Evaluation Inventory.....	224
G. Codes, Themes, Examples, and Frequencies for Evaluator Attribute Item.....	234
H. Codebook Development.....	235

**Tables**

1.01	Teacher evaluation status summaries for target states.....	32
3.01	States awarded Race to the Top grant funding in the first two rounds.....	73
3.02	Total MEA members that are K-12 music educators by state.....	84
4.01	Means and standard deviations for evaluation purpose.....	105
4.02	Paired samples <i>t</i> -test for evaluation purpose.....	106
4.03	Means and standard deviations for measures of teacher performance.....	108
4.04	Paired samples <i>t</i> -test for measures of teacher performance.....	109
4.05	Frequencies for type of student achievement data considered in music teacher evaluation.....	111
4.06	Frequencies for people involved in music teacher evaluation.....	112
4.07	Means and standard deviations for participants' beliefs about primary evaluator's competence.....	114
4.08	Means and standard deviations for participants' overall impressions of the evaluation process.....	117
4.09	Means and standard deviations for participants' overall impressions of evaluation benefits.....	119
4.10	Means and standard deviations for participants' overall impressions of evaluation impact.....	120
4.11	Pearson correlation matrix for facets of music educators' perceptions of teacher evaluation and total years teaching experience.....	122
4.12	MANOVA summary table for effect of individual differences upon upon teacher perceptions.....	126
4.13	Follow-up ANOVA summary table for individual differences on teacher perceptions.....	124
4.14	Scheffe post-hoc results for perceptions of evaluation benefits.....	125
4.15	Codes for responses reflecting "impact" theme.....	131

4.16	Codes for responses reflecting “concerns” theme.....	136
4.17	Codes for responses reflecting “teacher reaction” theme.....	140
4.18	Codes for responses reflecting “impact” theme: professional status.....	145
4.21	Codes for responses reflecting “concerns” theme: professional status.....	151
4.22	Codes for responses reflecting “teacher reaction” theme: professional status.....	153

**Figures**

1.01	Theory of action underlying teacher evaluation and school improvement.....	12
3.01	Strengths and weaknesses of online survey research.....	80
4.01	Type of school in which study participants taught music.....	101
4.02	Participants' primary teaching level.....	102
4.03	Participants' primary area of teaching.....	103
4.04	Evaluator with the most authority in music teacher evaluation.....	113

## Chapter I

### Introduction

*“Probably no other profession is subjected to more assessments with less effect than is the teaching profession...One cannot help but ask whether, and to what extent, these various assessments are contributing to improvements in the quality of any individual teachers or to the quality of the workforce as a whole.” (Kennedy, 2010, p. 1)*

### The Accountability Movement

Over the past several decades, an intense focus on accountability has characterized the United States education system. According to Rothman (1995), as cited by Leithwood and Earl (2000), educational accountability is defined as “the process[es] by which school districts and states attempt to ensure that schools and school systems meet their goals” (p. 2). The articulation of accountability standards, coupled with the use of performance-based incentives and sanctions, is viewed as a critical mechanism for improving schools and the education system overall.

In recent years, economic conceptualizations of educational accountability have been predominant, emphasizing the notion that “teachers must exhibit some standard of competency or performance, and schools must devise methods of relating expenditures to outcomes” (Ornstein, 1988, p. 12). In the current era of *test-based* accountability, school performance and teacher effectiveness are often linked to student achievement, as measured by standardized assessments (Figlio & Loeb, 2011). This concept, known as performance reporting, presumes that “information on educational outcomes is necessary to enable constituents to appraise the proficiencies of the schools” (Levin, 1974, p. 364). Because current conceptualizations of performance reporting center on *test-based* accountability, many perceive *accountability* and

*testing* to be one-in-the-same. In fact, the use of testing in United States schools far predates current perceptions of accountability.

**History of testing and accountability in the United States.** As early as the 19<sup>th</sup> century, schools used specific examinations to gauge student learning in various curricular areas and to inform administrative policy decisions. Notably, achievement was viewed more as a byproduct of individual student ability and effort, rather than the responsibility of the teacher or school (Cuban, 2004; Lavigne & Good, 2014; Ravitch, 2002). Teachers themselves were required to pass content exams and interviews to determine their readiness for the classroom, but once employed, they were no longer subject to tests of their professional competence.

Educational psychologist Edward L. Thorndike is commonly credited with launching and formalizing the United States school testing movement in the early 20<sup>th</sup> century. Thorndike worked toward perfecting and standardizing educational testing to enhance professional practice, and by the 1920s there were dozens of tests developed to measure achievement in major elementary and secondary school disciplines. Thorndike did not intend for educational testing to be used as a means of increasing the public's control of schools or to inform overall accountability; rather, Thorndike was hopeful that a scientific approach to education might allow schools to operate more independently and experience less public scrutiny (Ravitch, 2002).

The testing movement evolved with the growth of progressive education throughout the 1930s and 40s. Progressivists, who tended to reject many precepts underlying school testing initiatives, *did* support using test scores to place students (with varying levels of academic interest and commitment) in courses or activities best designed to support their learning. Despite testing being a growing part of education, there was no widespread support for using students' test scores as a mechanism to hold teachers or schools accountable. However, the increasingly

prevalent practice of social promotion (i.e., advancing students to the subsequent grade or level regardless of achievement, in order to protect their psychological well-being), created a perception that students were less responsible for their own achievement (Ravitch, 2002) and set the stage for a paradigm shift surrounding exactly *who* should be held accountable in education.

In 1965, President Lyndon Johnson initiated efforts to address poverty issues in the United States' by focusing on public education. The ensuing legislation, known as the Elementary and Secondary Education Act (ESEA), aimed to “strengthen and improve the educational quality and educational opportunities in the Nation’s elementary and secondary schools” (United States 89<sup>th</sup> Congress, 1965). Sociologist James Coleman, who subsequently was commissioned by the United States Department of Health, Education and Welfare to explore factors that determine educational outcomes, released his report entitled “Equality of Educational Opportunity” in 1966. Among Coleman’s major conclusions was the notion that school funding levels do not significantly affect student achievement, but student background and teacher effectiveness *do*. The differences highlighted in the report motivated policymakers to remedy achievement gaps, and more testing data were required to determine whether such gaps were being closed. With the establishment of the National Assessment of Educational Progress (NAEP) in 1970, test scores were made increasingly available for public distribution, further fueling stakeholders’ focus upon student achievement (Lavigne & Good, 2014; Ravitch, 2002).

In 1983, the Secretary of Education, T.H. Bell, commissioned a report entitled, “A Nation At Risk: The Imperative for Educational Reform.” The report, which characterized United States students’ academic performance as a “rising tide of mediocrity” (National Commission on Excellence in Education, 1983), triggered the most recent era of educational accountability (Graham, 2005). To ensure the United States stayed competitive with other countries in the

global marketplace, accountability advocates emphasized the importance of educational quality and academic excellence (Demarest, 2010; Kaestle & Lodewick, 2007; Spring, 2011), theorizing that improving teacher quality was necessary to realize gains in student achievement.

By the early 1980s, education accounted for nearly 40% of every state's budget. Many state officials recruited business leaders for insights on how to better manage school organizations while improving student achievement. These business leaders, in turn, expected schools to be held accountable for their performance. Since that time, education has been characterized by a clash of competing paradigms: the first – the professional educator paradigm – which strives to insulate the profession from public pressure for accountability, focusing instead upon improving educational inputs (such as teacher training); and the second – the policymaker paradigm – which emphasizes performance-based incentives and sanctions as a means of improving both teacher performance and student achievement (Ravitch, 2002).

In 2001, President George W. Bush reauthorized the ESEA, renaming the legislation, “The No Child Left Behind Act” (NCLB). This iteration of the ESEA shifted the intent of the 1965 legislation, which focused upon a specific group of students (minorities and the impoverished), to instead encompass *all* students in *all* schools nationwide, thereby establishing a uniform set of expectations for every child that would effectively close achievement gaps (Spring, 2011). Through NCLB, school districts and states were required to provide annual reports of student proficiency, serving as evidence of their accountability. Those schools failing to make adequate progress toward student proficiency expectations across multiple years were subject to corrective actions, including the replacement of school faculty and administration. In addition to high standards for student performance, NCLB mandated high standards for teacher credentialing. NCLB required all teachers be *highly qualified*, meaning that teachers needed a

bachelor's degree and formal certification in the subject area in which they taught. This mandate was based upon the rationale that qualified teachers were more effective, and would thus contribute to greater gains in student achievement (Rockhoff, 2004; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997). It also was presumed that teacher credentials were an adequate proxy for teacher quality or effectiveness (Demarest, 2010). This new focus on educators reflected a belief that teacher quality, beyond other within-school or out-of-school factors, was the key determinant of student academic outcomes. Consequently, a new initiative toward teacher quality was born (Graham, 2005), and with that initiative, a new focus on teacher evaluation.

**Race to the Top.** Competitive grants awarded through the United States Department of Education's Race to the Top (RttT) Fund exemplify an economics-based approach to improve teacher effectiveness. Introduced by the Obama administration in 2009, RttT represents the most recent initiative intended to improve educational outcomes in the United States. The RttT grants were made available to state education agencies that followed reform guidelines specified by the federal government. States interested in applying for RttT grants were required to present a formalized plan demonstrating strategies for adopting Common Core standards into the curriculum, closing the achievement gap, developing high quality standardized assessments, and improving teacher effectiveness.

The Department of Education created a points system to help gauge the overall quality of each state's reform plan. Points reflected the relative weight or importance of various planning areas (e.g., "Data Systems to Support Instruction" and "Turning Around the Lowest Achieving Schools"), highlighting priorities for state-level reform. In applying for RttT funding, states developed strategies addressing each of these planning areas, and the Department of Education

assessed the quality of those plans by awarding an appropriate amount of points out of 500 possible. States with the highest total points across all six planning areas were rewarded with federal funding to help the plan for reform become a reality (United States Department of Education, 2010).

Of all the planning areas outlined by RttT, “Great Teachers and Leaders” received the strongest weight, with 138 of the total possible 500 points attributed to plans for improving teacher and administrator effectiveness (United States Department of Education, 2010). The emphasis on improved teacher effectiveness stemmed from research-based evidence connecting teacher effectiveness and student achievement (Rockhoff, 2004; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997). The explicit goal of this RttT initiative was to improve the effectiveness of teachers and administrators by creating more rigorous evaluation procedures.

RttT’s proposed evaluation system was multi-dimensional. Teacher and administrator effectiveness was to be determined, in substantial part, by student growth – primarily as reflected in student performance on standardized assessments. Recognizing the diverse dimensions of teacher quality, the “Great Teachers and Leaders” initiative also supported the consideration of classroom observations, student perception surveys regarding teacher performance, and portfolio artifacts representative of teacher impact in the classroom. RttT teacher evaluation guidelines specified that these multiple measures of teacher effectiveness were to be used to produce evaluative outcomes that determined how incentives and sanctions (appropriate to each teacher’s performance) were administered. To be eligible for RttT funds, interested states needed to develop a rigorous system for teacher evaluation that adhered to these guidelines (United States Department of Education, 2010).

In the first round of the RttT competition, a total of 40 states plus the District of

Columbia applied for federal funding by the initial January 2010 deadline. Of the applicants, a total of 16 states were selected as finalists. In the end, only two states – Delaware and Tennessee – were awarded RttT funding. Delaware’s reform plan earned 454 points out of a possible 500, and the state was awarded \$100 million in funding toward implementation. Tennessee’s plan earned 444 points and \$500 million of funding, reflecting that state’s larger population of elementary and secondary school students. A total of 14 other states were finalists in the first round, but did not receive funding for the reform proposal submitted. A second round of applications were accepted in June 2010, and of the second round applicants, nine more states plus the District of Columbia were awarded RttT funding. Among the second round winners were Rhode Island, North Carolina, and Florida.<sup>1</sup>

*Section summary.* As evidenced by initiatives such as NCLB and RttT, the policymaker paradigm has gained much traction among stakeholders in recent years. The professional educator paradigm, with its focus on inputs (i.e., funding and teacher training) rather than outputs (i.e., measurable performance achievement), can be perceived as a bureaucratic obstacle to reform. Accountability practices, as they are defined today, shift the national focus toward quantifiable products of the public’s educational investments. In contrast to early 20<sup>th</sup> century mores, the responsibility for attaining satisfactory growth and achievement is now placed largely on the teacher and school, rather than on the student (Ravitch, 2002). This responsibility shift is rooted in a belief that teacher quality drives student growth and academic success (Rockhoff, 2004; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997).<sup>2</sup> The recent wave of federally

---

<sup>1</sup> Some states that unsuccessfully applied for RttT funding, such as Colorado, still decided to implement RttT-inspired education reforms.

<sup>2</sup> While teacher effectiveness can as be empirically linked to student achievement, research suggests that non-school factors, such individual and family characteristics, have a much larger impact upon student achievement than teacher quality. School-related factors typically account for less than 30% of student’s success, while other variables, such as socioeconomic status, play a much larger role in student achievement (Berliner & Glass, 2014).

driven teacher evaluation systems emphasizes performance-based incentives and sanctions, largely in an effort to more effectively hold *teachers* accountable for student achievement and school improvement.

### **Teacher Evaluation: Conflicts and Concerns**

The teacher evaluation process may be used to (a) make determinations about teacher effectiveness (summative purposes), and (b) inform approaches to professional development (formative purposes). Summative evaluation purposes align with the output-oriented policymaker paradigm, while formative evaluation purposes reflect the professional educator paradigm. Marzano (2012) argues that evaluation systems designed to measure teacher effectiveness are distinctly different than those intended to foster professional development, but suggests it is important to employ an evaluation system that features both accountability and growth components. Goe, Biggers, and Croft (2012) indicate that, within the context of RttT funding and teacher evaluation reform, many schools emphasize summative objectives *over* formative objectives in the teacher evaluation process. Because teacher evaluations designed to promote professional development typically yield more information-rich data and feedback (Goe, Bell, & Little, 2008), evaluations conducted for high-stakes, summative purposes may inhibit teacher efforts to improve professional practice.

Beyond concerns regarding adequate professional development opportunities, summative-oriented teacher evaluations driven by RttT raise additional concerns about the use and interpretation of student growth data. While students' academic growth has long been a focus of stakeholders, considering such data in determining teachers' overall effectiveness is a recent phenomenon. RttT grant recipients were tasked with ensuring that student growth measures used to determine teacher effectiveness were (a) rigorous in expectations, (b) drawn

from two points in time, and (c) conducive to cross-classroom comparisons (Goe & Holdheide, 2011). To achieve these requirements, student growth data are most commonly derived from standardized test scores. However, the majority of educators (69%) teach in subject areas or grades for which there *are no standardized assessments*. According to Goe and Holdheide, non-tested subjects might include:

1. Subjects with standards that cannot be adequately or completely measured with a paper-and-pencil test (e.g., art, music, industrial arts, drama, dance)
2. Subjects in lower elementary grades for which students cannot be reliably tested with a paper-and-pencil population test (e.g., grades K–2)
3. Subjects/grades for which states have chosen not to test because of cost and priority relative to ‘core’ academic subjects (p. 2)

Some researchers question the validity of the assessments used to determine student growth (Baker, Oluwole, & Green, 2013; Hill, 2009). These concerns are amplified when teachers in non-tested subjects are evaluated using school-wide growth data that may not be attributable to the individual teacher or reflect student learning in a non-tested content area (Goe & Holdheide, 2011). Given the high-stakes nature of incentives and sanctions associated with RttT-driven teacher evaluation policies, all measures of teacher effectiveness may not be not equally valid representations of each teacher’s unique contributions to student learning in their content area of expertise.

### **Research Problem**

There is a long history of testing and evaluation in U.S. schools. In recent years,

education reforms have shifted the accountability for achievement outcomes from students to teachers. Because many view teacher quality as a major determinant of student learning, there has been a push for more robust and rigorous teacher evaluation systems. An economics-based approach to teacher evaluation has been advanced by federal policies and funding in the form of Race to the Top (RttT) grants. RttT-driven teacher evaluation systems, however, appear to emphasize summative evaluation functions and teacher accountability over formative evaluation functions and teacher development. Moreover, major concerns have arisen concerning the use of standardized test scores and school-wide growth measures to determine teacher effectiveness.

Most analyses of teacher evaluation policy fail to consider questions of broad applicability and veracity where teachers working in subjects other than reading, writing, and math are concerned. Music teachers, in particular, may be at risk, given the unique challenges associated with evaluating music teaching and learning in K–12 schools. Evaluation processes are typically designed for the traditional classroom teacher, and as such, may not fit or adapt well to the performance-based, expressive environments characteristic of the music classroom. Maranzano (2002) points out, “Educational administrators charged with the responsibility for evaluating school personnel confront an additional challenge when compelled to apply general models that may not be suited to the highly complex world of performing arts instruction” (p. 6). Indeed, music educators have historically struggled to find meaning in evaluations conducted by school administrators who lack experience with or understanding of artistic disciplines (e.g., Brophy, 1995; Shaw, 2013; Taebel, 1990a). Evaluation systems emphasizing high-stakes, summative outcomes may further draw evaluators’ attention from the process of providing music educators’ practical guidance toward meaningful professional growth.

In addition to professional development concerns, music educators may be faced with the

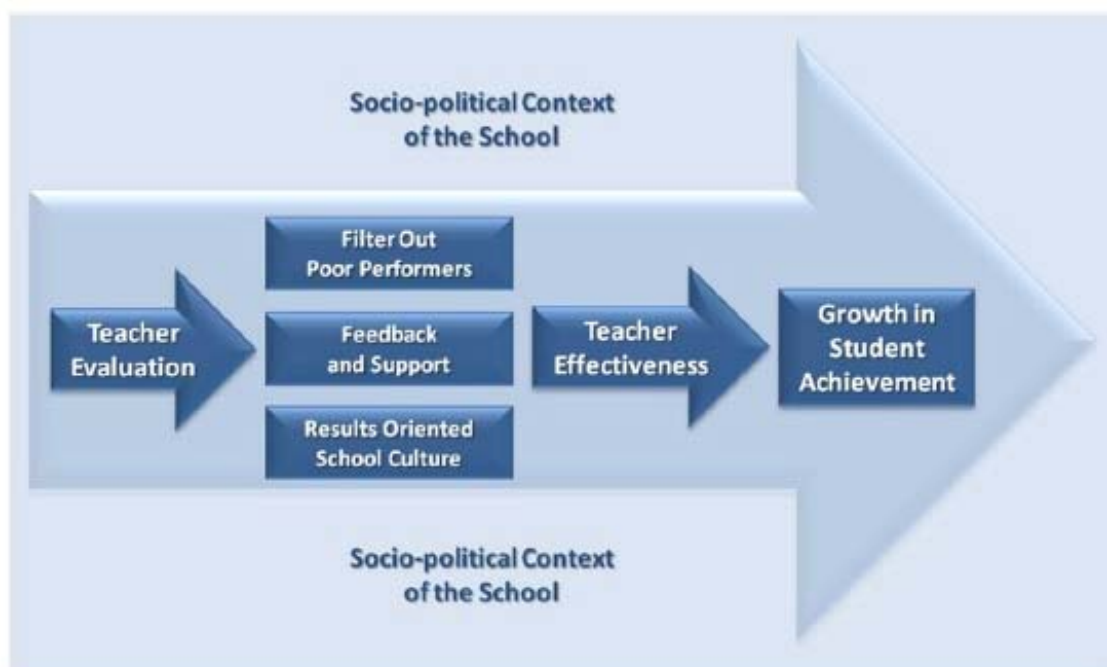
challenge of having a significant portion of their overall evaluation predicated upon invalid or unreliable student growth measures. With high-stakes incentives and sanctions associated with RttT-driven teacher evaluations, it is critical that student growth measures considered in determining music teacher effectiveness validly assess *music* student learning, drawing direct connections to the music teacher's contributions to that learning. Basing music teacher effectiveness on invalid or unreliable measures of music student growth may negatively impact music educators' feelings of professional worth and overall commitment to the profession.

Oftentimes, when examining a new policy's effectiveness, analysts fail to explore the perspectives and experiences of all parties involved (e.g., teachers, administrators, school board members, students), relying instead upon narrowly defined outcomes (i.e., student test scores) to determine a given policy's success (Darling-Hammond, 1990a). The more teachers see evaluations as being connected to and having a positive impact on their teaching practice, the more likely they are to support new evaluation policies (Milanowski & Heneman, 2001; O'Pry & Schumacher, 2012; Schumacher, 2010). In the face of new policies and reforms, teacher perceptions regarding the instrumentality (i.e., usefulness), congruence (i.e., relevance), cost (particularly in terms of time), difficulty of use (i.e., ease of applicability), and value of proposed policies will ultimately determine teacher investment in policy implementation (Doyle & Ponder, 1977; Mohlman, Coladarci, & Gage, 1982). Currently, limited research focuses on music teachers' perceptions of recent teacher evaluation initiatives, and no known studies examine the experiences of music educators across multiple states implementing RttT-inspired evaluation processes. Examining the unique challenges of music teacher evaluation from the perspectives of music educators themselves could help clarify how RttT policy implementation impacts the instructional decisions, motivations, professional commitment, and career satisfaction of teachers

within and outside of music.

### Teacher Evaluation: Theories of Action and Motivation

Teacher evaluation policies, such as those outlined by RttT, characteristically assume a direct, linear relationship between teacher quality and student achievement. Hallinger, Heck, and Murphy (2014) outline this logic-based theory underlying teacher evaluation and school improvement, which has dominated evaluation discourse throughout the past decade and is largely supported by advocates of recent teacher evaluation policy reform (See Figure 1.01).



**Figure 1.01.** Theory of action underlying teacher evaluation and school improvement. Reprinted from “Teacher Evaluation and School Improvement: An Analysis of the Evidence,” by P. Hallinger, R.H. Heck, and J. Murphy, 2014, *Educational Assessment, Evaluation, and Accountability*, 26, p. 8. Copyright 2014 Springer. Reprinted with permission.\*

\*Springer and *Educational Assessment, Evaluation and Accountability*, 26 (2014), p. 8 in Teacher Evaluation and School Improvement: An Analysis of the Evidence,” by P. Hallinger, R.H. Heck, and J. Murphy, Figure 1.01 Theory of action underlying teacher evaluation and school improvement is given to the publisher in which the material was originally published with kind permission from Springer Science and Business Media.

The model shown in Figure 1.01 is “predicated on the strength of the causal relationship between teacher quality and growth in student learning” (p. 7). According to the model, teacher performance evaluations contribute to student achievement by effectively filtering out those educators who fail to have a consistent, positive impact upon student learning. The model also presumes evaluations provide teachers meaningful feedback that helps improve their overall instructional quality. Furthermore, this model purports to develop a “results-oriented school culture” (p. 8) that fosters openness to new policies, which may further enhance teacher effectiveness. While the model outlined in Figure 1.01 may make sense from a managerial standpoint, there is little empirical evidence supporting the model’s premise that teacher evaluation itself drives growth in student achievement through a combination of direct and indirect effects.

Hallinger et al. (2014) attribute limitations associated with current school improvement models to policymakers’ assumptions regarding the validity and reliability of tools used to gauge teacher effectiveness, particularly those emphasizing student achievement data. Citing the 1966 Coleman Report, the authors point out that many factors, beyond the quality of the teacher, inform student performance, recalling, “It was previously deemed unfair to hold teachers accountable for a process-product relationship between teaching and learning that was determined to such a large degree by factors outside the control of the individual teacher” (p. 9). Ravitch (2010) agrees, cautioning that when teachers are “filtered out” due to a perceived lack of impact upon student learning, such decisions are often made based upon student test scores alone, which is inadequate evidence to support such sanctions. She goes on to argue:

The trouble with test-based accountability is that it imposes serious consequences on children, educators, and schools on the basis of scores that may reflect measurement error, statistical error, random variation, or a host of environmental factors or student attributes. None of us would want to be evaluated – with our reputation and livelihood on the line – solely on the basis of an instrument that is prone to error and ambiguity. (p. 166)

According to Hallinger and colleagues, proponents of robust teacher evaluation models believe the outcomes are worth the “considerable investment of human and financial resources required for its effective implementation” (p. 9). Although teachers do have some effect on student learning, the degree of that effect may be overestimated, the manner in which we measure that effect may be unreliable, and the ensuing consequences for the teaching profession put additional stress on already limited resources.

Ravitch (2010) asserts that contemporary teacher evaluation models, such as the model described by Hallinger et al. (2014), are based on notions of “punitive accountability” (i.e., low scores used for reprimand) over “positive accountability” (i.e., low scores used to promote positive change) (p. 163). In other words, current evaluation models center on the philosophy that *sanctions motivate teachers*. The concept that rewards (e.g., money, professional status) or sanctions extrinsic to the task of teaching can serve as motivation to improve teacher performance, thereby improving student achievement, is central to external motivation theory (Firestone, 2014). Within this framework, stakeholders stress the importance of quantitative data (such as student test scores) in informing the distribution of various rewards and incentives.

Experts in economics, organizational studies, and political science have developed elaborate incentive systems in an effort to shape various behaviors measured in relation to a variety of contexts and problems. The major dilemma facing policymakers and administrators who implement such systems within *educational* contexts, is that the types of incentives teachers value may differ from those posited by economic theorists. Furthermore, while teachers may produce outcomes that qualify them for such incentives, the methods used to achieve certain outcomes may not always be educationally valid.

According to Firestone, teacher evaluation should emphasize *intrinsic* incentives, where “the activity [of teaching] is so interesting that no additional incentive is needed” (p. 2). This approach is heavily rooted in ideas of professional development, driven by teachers’ feelings of self-efficacy and autonomy in the classroom, so that teachers might “work for personal interest without oversight or coercion” (p. 2). Wise and Darling-Hammond (1985) refer to this as the “professional conception” (p. 30) of teacher evaluation, outlining the following formula embodying this approach:

1. Involves the teachers in the development and operation of the teacher evaluation process;
2. Bases evaluation on professional standards of practice that are client-oriented;
3. Recognizes multiple teaching strategies and learning outcomes;
4. Treats teachers differently according to their teaching assignments, stages of development, and classroom goals.

(p. 30)

Firestone (2014) argues that the emphasis on extrinsic motivation in current teacher evaluation policy leaves little to no room for teachers who embody intrinsic motivation and identify with the professional conception of teaching. In fact, rewards and sanctions may effectively undermine motivation to seek intrinsic rewards. For example, while evaluations require teachers to be transparent about their performance (i.e., be forthcoming about strengths and weaknesses) to facilitate meaningful growth and development (intrinsic reward), in extrinsically oriented systems, such transparency could lead to undesirable professional consequences, including dismissal. High-stakes professional consequences may therefore inhibit many teachers from fully committing to the evaluation process.

Firestone also observes that the time needed to achieve extrinsically oriented outcomes competes with time necessary for teachers to establish intrinsically oriented motivations. In other words, the high-stakes nature of extrinsic consequences, such as merit pay and professional status, may draw teachers' attention away from tasks associated with professional growth and development. Finally, extrinsic-based evaluations systems do not typically provide teachers with the quantity or quality of feedback necessary for meaningful professional development. Over time, this may result in a diminished sense of professionalism and a weakened commitment to teaching.

### **The Status of Music Education in the United States**

To more thoroughly situate these conflicts in context of music teaching, it is important to understand a general profile of music education positions in the United States. In 2012, the National Center for Education Statistics (NCES) released a report on the status of arts education in the United States. The purpose of this national survey study was to collect information regarding the extent of arts instruction and resources available for arts education, among other

factors. One dimension of this report specifically explored music education. Participants in the survey study included building principals as well as elementary and secondary music teachers. While principals were surveyed on various characteristics of their school music programs, music teachers answered questions on topics specific to the classroom, such as assessment practices and teaching responsibilities.

Survey results revealed that over half (54%) of elementary music teachers taught in more than one building, while nearly one-quarter (24%) of teachers taught music classes outside normal school hours. At the secondary level, over one-third of music educators taught classes outside the normal school day, and another 46% of music educators taught in multiple schools (National Center for Education Statistics, 2012). Typically, traditional classroom teachers are committed to a single school. Music teachers committed to multiple buildings must consider different administrators' diverse professional expectations. Furthermore, among traditional classroom teachers, there are more limited expectations to teach formal courses outside normal school hours. The amount of required instructional obligations outside the typical school day further distinguishes music educators from traditional classroom teachers.

In an earlier study, Gardner (2010) also determined music teachers' job attributes differed significantly from other classroom teachers, including hours worked per week and type of teaching assignment (i.e., full- versus part-time or itinerant employment). Music teachers were more likely to hold part-time or itinerant positions, and they reported receiving less classroom support for students with individualized education plans. Gardner determined that these and other job attributes, together with teacher attributes (i.e., age, years of teaching experiences) and perceptions of the workplace (i.e., perceived support from administration, autonomy in instructional practice) contributed to music teachers' overall job satisfaction and teacher status

(in terms of retention, turnover, and attrition). With these critical professional consequences at stake, policymakers must take into account the unique nature of music teaching and learning when designing evaluative processes.

**The challenge of evaluating music teachers.** Barrett (2011) states, “Quality in music teaching is a complex and sophisticated notion not easily captured by platitudes and checklists” (p. 1). This statement reflects the experiences of many music teachers who find their evaluations to be lacking in accuracy or complexity. While many qualities of effective teachers certainly overlap with necessary skills for effective *music* teaching, Brophy (1995) raises the question:

Is it appropriate to evaluate music educators exclusively on general competencies, or does effective music teaching involve certain teaching behaviors, characteristics, and attributes that are so significantly unique that they demand their own set of evaluative criteria? (p. 1)

Key concerns surrounding music teacher evaluation typically include the fit of the evaluative instrument (in particular, that which is used in classroom observation settings) and the qualifications of the evaluator. Efforts to address these concerns are evident in the development of classroom observation instruments specific to music teaching, such as the teacher evaluation workbook (*Workbook for Building and Evaluating Effective Education in Music*) published by the National Association for Music Education (NAfME) in 2013. Other concerns include the overall purpose or objective of the evaluation process, as well as teachers’ perceptions of evaluation in terms of clarity and equity.

**Evaluative instrument.** Maranzano (2000) suggests that traditional evaluation methods

have “limited applicability” (p. 267) to performing arts educators. Despite the common behaviors that are integral to effective teaching in all subject areas (e.g., effective planning, accommodating needs of various learners), effective music teachers exhibit specialized behaviors that may not be recognized by many generic evaluation systems. In particular, the classroom observation component of teacher evaluations can be problematic, as there are many teaching behaviors unique to the music classroom. Maranzano points out the “split-second analysis, gestures, pacing, and pedagogical communication” (p. 268) evident in the music classroom represent behaviors that traditional classroom observation instruments may not effectively measure. Oftentimes music teachers make on-the-spot aural discriminations in response to student performance, and these diagnostic-prescriptive sequences inform subsequent instruction or actions. Music educators also engage in more non-verbal communication and instruction (Gipson, 1978; Kurkul, 2007) than do typical classroom teachers, further highlighting the unique nature of music teaching and the subsequent challenges that arise when evaluating music educators.

Many researchers echo the concern that there is a need for content-specific evaluation tools suitable for evaluating *music* teaching, particularly through classroom observations (Brophy, 1995; Collins, 1996; Colwell, 2010; McAllister, 2008). Taebel (1990b) proposed that music teacher observation instruments should highlight skills such as modeling and accuracy of error detection, and these instruments should be designed to recognize both verbalized and performance-based student engagement. Without such considerations, the generalized nature of traditional teaching evaluation models may yield biased results when used to assess music teacher effectiveness (Brophy, 1995; Maranzano, 2000).

To help account for diverse practices in the classroom, current teacher evaluation systems

typically embrace a multi-dimensional approach to assessing teacher effectiveness, taking various measures into consideration. In addition to classroom observations, overall evaluations also consider measures such as student achievement/growth, student input or feedback, and participation in professional development activities. While including other dimensions of teaching effectiveness can paint a more comprehensive picture of music teachers' overall impact in the classroom, it is critical to examine how these measures specifically apply to music teaching. Music is an inherently aural phenomenon, and music teachers – as well as musicians – must make many decisions based upon in-the-moment aural judgments. When evaluating a music teacher in a classroom observation setting, an evaluator might struggle to identify and assess these aurally based decisions. Similarly, because many music classrooms are performance-based, student achievement or growth may be demonstrated through singing, playing an instrument, or creating music in another manner. How might an evaluation system account for these unique displays of student knowledge, and how might the evaluator identify growth? While implementing a multiple-measures approach to teacher evaluation enhances its usefulness, it is still important for stakeholders to carefully consider how well such systems align with the unique circumstances that characterize music teaching.

*Evaluator expertise.* Though the appropriateness of the evaluative instrument is critical, Goe and colleagues (2008) assert that a valid evaluative instrument alone does not guarantee classroom observation evaluation data are credible. An evaluator's training, expertise, and interpretation of instructions and instrument descriptors also will determine whether the evaluative process is reliable and accurate. Additionally, teachers are more likely to favor evaluation processes facilitated by a capable, objective evaluator (Milanowski & Heneman, 2001; O'Pry & Schumacher, 2012). Without appropriate evaluator training or expertise,

evaluative ratings may be inaccurate or prone to misinterpretation, and the usefulness of observational data can be compromised (Duke, 1995).

Among music teachers specifically, several researchers have determined the importance of musically trained evaluators in accurately assessing classroom performance (Csipkes, 2011; Duke & Blackman, 1991; Goddard, 2004; Schmidt, 1992), while other researchers have shown that, even among musically trained evaluators, there can be distinct disagreements regarding a music teacher's classroom performance (Delaney, 2007; Finnegan, 2003). While it is clear that an evaluator's content-area expertise plays some role in music teacher evaluations, no known research provides a multi-state summary of music teachers' perceptions of evaluator competence. By exploring the relationship between music teachers and their evaluators, stakeholders can gain a more comprehensive understanding of how to ensure music teacher evaluations are trustworthy and meaningful.

### **Music Education's Response to Evaluation Initiatives**

Doerksen (2006) provided a specific set of guidelines for appropriately appraising music teachers in his book *Evaluating Teachers of Music Performance Groups*. In this text, Doerksen suggested that proper training and guidance could prepare any educator to effectively evaluate music teachers. For many years, this publication served as the primary effort to encourage accurate evaluations of music educators. More recently, in September 2011, NAFME developed a position statement on music teacher evaluation. In this statement, NAFME recommends that music teacher evaluations be based on multiple measures defining the types of achievement data that are appropriate for evaluating music teachers. For example, NAFME asserts student achievement data must be "directly attributable to the individual teacher, in the subject area taught by that teacher" (p. 1), cautioning against using school-wide data to evaluate music

teachers' performance. The organization points out the importance of considering music student achievement in the context of class size and instructional time and advocates for musically trained evaluators in classroom-based observations (National Association for Music Education, 2011<sup>a</sup>).

In addition to recommendations for teacher evaluation practices, NAFME compiled an advocacy guide for practicing teachers, focusing upon strategies for managing emerging teacher evaluation practices. This guide includes an outline of the NAFME evaluation recommendations and provides practicing teachers with suggestions for addressing evaluation practices with building administrators. Music teachers also are encouraged to take an active role in understanding evaluation practices by participating in local and state level discussions, as well as allying with colleagues to advocate for sound approaches to music teacher evaluation at the local level (National Association for Music Education, 2011<sup>b</sup>).

The Society for Music Teacher Education (SMTE), another prominent national music education organization, established an Area of Strategic Planning and Action (ASPA) group focused specifically on music teacher evaluation. This group was created to assist practicing music teachers and researchers in becoming familiar with policies and trends relevant to music teacher evaluation. Additionally, the ASPA highlights recent research relevant to music teacher evaluation. The group also works toward gaining a more comprehensive understanding of music teacher evaluation practices among different states (Society for Music Teacher Education, 2013).

Following suit, many state-level organizations have made efforts to address music teacher evaluation practices, assembling task forces devoted to relevant policy changes at the state level. For example, the Ohio Music Educators Association (OMEA) and the Ohio Alliance for Arts Education (OAAE) partnered to develop a position statement on essential elements of

music teacher evaluation, while also exploring state-level student growth measures in music (Ohio Music Educators Association, 2012). In Michigan, The Partnership for Music Education Policy Development (PMEPD) issued a statement describing who should be evaluating music teachers and how those teachers should be evaluated (Partnership for Music Education Policy Development, 2012). In particular, PMEPD called for musically trained specialists with successful classroom teaching experience to evaluate music instruction and classroom materials. Building administrators should be responsible for evaluating other aspects of music teachers' professional performance, including collaborative and communicative skills. PMEPD also offered suggestions as to which measures should comprise music teachers' overall evaluation.

Music education researchers have also brought attention to teacher evaluation in recent issues of journals specific to music and arts education. In September of 2014, Music Educators Journal published several articles on teacher evaluation (e.g., Nielsen, 2014; Overland, 2014). Also in the fall of 2014, Arts Education Policy Review featured articles specific to teacher evaluation, both at the preservice level (Peterson, 2014) and among K–12 teachers impacted by Race to the Top (Aguilar & Kapalka Richerme, 2014). Recent attention to music teacher evaluation in these prominent journals indicates this issue is a top priority among music educators. Despite the growing attention to the issues surrounding music teacher evaluation, there remains little consensus as to exactly how music educators should be evaluated. Moreover, little is known about music teachers' perceptions of emergent evaluation systems and their perceived impact upon teacher quality and student learning.

### **State-Level Approaches to Evaluation**

Among the states earning RttT grant funding, there exists some variation as to how teacher evaluation is executed. Though each state's policy falls within the prescribed guidelines

set forth by the RttT program, evaluative approaches may be customized to the specific needs and goals of each state. As states near the end of the four-year grant period, stakeholders continue to adapt and refine teacher evaluation policies to address teacher concerns, utility, and overall practicality. At the time of data collection for the current study, each of these four states was currently in the second or third year of implementation.

**Tennessee.** Tennessee was one of the first recipients of RttT grant funding, earning \$500 million toward reform initiatives in March 2010. Tennessee began full implementation of its new teacher evaluation systems during the 2011–2012 school year. Their evaluation system, the Tennessee Educator Acceleration Model (TEAM), is “a combination of frequent observation, constructive feedback, student data, and meaningful professional development” (Tennessee Department of Education 2013). The TEAM system is a presumptive state evaluation model, through which districts have the option to opt-out in favor of an amended version that adheres to specific guidelines.

To specifically address the student growth component of teacher evaluation, Tennessee developed the Tennessee Value-Added Assessment System (TVAAS), a longitudinal data-based process for determining students’ academic growth over time. TVAAS student growth data count toward 35% of a teacher’s overall effectiveness, while an additional 15% of student achievement data can be drawn from other sources. For teachers in tested subjects, TVAAS data is attributed to individual teachers, while teachers in non-tested subjects were initially evaluated based upon school-level composite TVAAS data (Tennessee Department of Education, 2013).

Because school-level composite TVAAS data were not considered a satisfactory means of validly assessing student growth in the fine arts, Tennessee explored alternative approaches to collecting value-added data for teachers in these content areas. In 2011, Tennessee developed a

system to pilot among selected arts teachers in the state. A scale model was piloted across three districts during the 2012–2013 school year. During the 2013–2014, Tennessee released its Fine Arts Growth Measure system (FAGM), a model designed specifically to measure student growth in the arts, and twelve districts implemented the model, including three of the largest school districts in Tennessee (Tennessee Department of Education, 2013).

This portfolio system requires arts teachers to collect five pieces of pre- and post-test evidence signifying student growth in their area. The teachers must collect this evidence from a representative sample of their students, and the artifacts need to encompass three of Tennessee's four arts learning domains (perform, create, respond, connect). Arts teachers self-rate their contribution to student growth based upon these artifacts, and a trained peer reviewer also assesses student growth based on the provided evidence. Though the teachers self-rate, the overall growth score is determined in the blind peer review. The growth evident in the FAGM portfolio system constitutes the required 35% value-added component of teachers' overall effectiveness. Even if the TVAAS school-level composite data regards an arts teacher as more effective than the FAGM portfolio system, arts teachers are required to submit FAGM data as their value-added evidence (Tennessee Fine Arts Student Growth Measures System, 2013). While the FAGM portfolio system involves content-area experts in the blind peer review, a building or district administrator (who may or may not be an expert in music) facilitates the remaining 50% of Tennessee music teachers' evaluation.

**North Carolina.** North Carolina received its RttT award in September of 2010. North Carolina also approached implementation gradually, with 2012–2013 serving as the first operational school year of their educational effectiveness model. In July 2011, North Carolina expanded its statewide teaching standards to encompass qualities and skill sets expected of

successful teachers in 21<sup>st</sup> century classrooms, and they include:

1. Teachers demonstrate leadership
2. Teachers establish a respectful environment for a diverse population of students
3. Teachers know the content they teach
4. Teachers facilitate learning for their students
5. Teachers reflect on their practice
6. Teachers contribute to the academic success of students

(Public Schools of North Carolina, 2013<sup>a</sup>)

North Carolina's 6<sup>th</sup> standard reflects the evaluation system component focusing upon student growth data. According to the Public Schools of North Carolina State Board of Education, a teacher's rating on this standard is "determined by student growth value as calculated by the statewide growth model for educator effectiveness" (Public Schools of North Carolina, 2013<sup>a</sup>, p. 8). The student growth value classifies teachers in one of three categories:

1. Does not meet expected growth: The student growth value for the teacher is lower than what was expected as per the statewide growth model.
2. Meets expected growth: The student growth value for the teacher is what was expected per the statewide growth model.
3. Exceeds expected growth: The student growth value for the teacher exceeds what was expected per the statewide growth model.

(Public Schools of North Carolina, 2013<sup>a</sup>)

A teacher's value-added classification is based on the growth data for those students taught by that teacher, when data are available. If no subject-specific growth data are available for a given teacher, building-level growth data are used for that teacher's evaluation. Recognizing that such an approach provides an inequitable evaluation for many teachers, North Carolina piloted a system in January 2013 entitled, "Analysis of Student Work" (ASW). This system serves to fulfill the sixth state requirement, "Teachers Contribute to the Academic Success of Students."

North Carolina's ASW approach is designed for teachers in non-tested subjects, such as music. The system requires those teachers to submit work samples from randomly selected students. Artifacts are compiled from three students in each class taught by that teacher, and these work samples are collected at the beginning and end of each course. Trained content area experts examine these artifacts in a blind review process in order to determine student growth from one point to the other. Based upon student growth determinations, the teacher earns an evaluation rating reflecting the amount of growth evident among students. Due to the success of the small-scale pilot in January 2013, North Carolina implemented a large-scale pilot test of this system in the spring of 2014. North Carolina educators hope to have this system fully implemented by the 2014–2015 school year (Public Schools of North Carolina, 2013<sup>b</sup>). Though the ASW approach involves content-area experts in the blind peer review, a building or district administrator executes the remainder of North Carolina music teachers' overall evaluation. Similar to Tennessee, this administrator may or may not have expertise in music.

**Rhode Island.** The Rhode Island Department of Education (RIDE) developed the Rhode Island Teacher Evaluation and Support System (RITESS) in 2011, and the state and pilot tested the system during the 2011–2012 school year. Based upon statewide feedback, RIDE revised the

RITESS for a June 2012 release, and by the 2012–2013 school year, the revised model was fully implemented, though the model underwent a third revision for the 2013–2014 school year. The Rhode Island Model is based upon multiple measures to gain a comprehensive perspective on teacher effectiveness. These measures include:

1. Professional Practice – A measure of effective instruction and classroom environment as defined in the Teacher Professional Practice Rubric
2. Professional Foundations – A measure of instructional planning and the contributions teachers make as members of their learning community as defined in the Teacher Professional Foundations Rubric
3. Student Learning – A measure of a teacher’s impact on student learning through demonstrated progress toward academic goals (Student Learning Objectives, with the Rhode Island Growth Model in tested grades and subjects)

(Rhode Island Department of Education, 2014, p. 4)

Among each of the categories listed, teachers earn a score, and the sum of these scores results in an overall effectiveness rating, ranging from ineffective to highly effective. Rather than defining a specific percentage by which student growth contributes to teachers’ overall effectiveness, Rhode Island created a system that weighs each component in a manner where student growth is considered the preponderant criteria. In other words, student growth is weighted most strongly in the evaluative process.

In the Rhode Island system, student growth is measured through Student Learning/Outcome Objectives (SLOs and SOOs) and by the Rhode Island Growth Model (RIGM). SLOs/SOOs are standards-aligned, measurable academic goals partnered with

meaningful assessments. Teachers design these goals, and they involve collecting baseline data to effectively track student growth over time. In Rhode Island, teachers must create a minimum of two SLOs per year, typically collaborating with other teachers in the same or similar content areas.

The RIGM is a more structured, statistical model that provides a supplementary perspective on student achievement. RIGM examines not only the growth evident in the SLOs/SOOs, but also student proficiency as measured through standardized tests, so as to determine a comprehensive profile of student achievement. The student proficiency data is obtained from student performance on the New England Common Assessment Program (NECAP), which focuses upon math and reading achievement for students in grades 3–7. Local Education Agencies (LEAs) determine which educators contributed to math and/or reading for students in these grades, and according to RIDE, those teachers considered “contributing educators” to these areas can include *all* educators, including music specialists. Using standardized assessment data from reading and math, median growth percentiles are calculated for each educator’s student roster, and these growth percentiles reflect the student proficiency component of the RIGM process. The 2013–2014 school year was the first year in which RIGM results were considered in those teachers’ overall effectiveness ratings (Rhode Island Department of Education, 2014).

**Florida.** In Florida, both RttT and the state’s Student Success Act impacted teacher evaluation reform. Though Florida applied for RttT grant funding in the first round, the state was not awarded any funds with their initial reform plan. As Florida developed a new submission, state teachers of non-tested expressed concern that the value-added component of their evaluations would be based upon content outside their area of specialty. As such, Florida’s

revised approach to teacher evaluation reform focused heavily upon developing assessments for content areas that were not measured by the Florida Comprehensive Assessment Test (FCAT).

Upon winning an RttT grant worth \$700 million dollars in the second round, Florida began full implementation of its new teacher evaluation systems during the 2011–2012 school year. The Florida Department of Education (FDOE) formed a new department entitled “Race to the Top Assessments.” This department managed LEA initiatives to create high-quality assessments for content areas outside of the FCAT, initiating a \$20 million dollar state-level grant process incentivizing LEAs to develop assessments for these areas (Florida Department of Education, 2012). From this process, the RttT Performing Fine Arts Assessment Project (PFAAP) was created.

PFAAP’s purpose is to “design and develop high-quality assessment items for dance, music, and theater ‘highest enrolled’ courses (i.e., those courses that have the most students statewide) in the state of Florida” by creating a standards-based item bank designed to assess individual student achievement in the performing arts. These items, written and reviewed by practicing teachers with three or more years of experience in the field, are currently being pilot-tested throughout the state, and the items address all three facets of performing, creating, and responding to music, as outlined by the Next Generation Sunshine State Standards. Florida hopes to have fully implemented an assessment process for the performing arts by the 2014–2015 school year. In Florida, at least 50% of teacher evaluations are based upon student growth as measured by statewide assessments, and these art-specific assessments are designed to fulfill that 50% requirement (Center for Fine Arts Education, 2013).

However, prior to the implementation of PFAAP, Florida intended for the student growth component of arts teacher evaluations to consist of students’ standardized test scores in other

disciplines. This approach also affected teachers in other non-tested subjects, such as biology, for whom state wide standardized test data were unavailable. In many cases, this involved teachers being evaluated using growth data from students they did not even teach. In the spring of 2013, the National Education Association (NEA) and Florida Education Association (FEA), together with seven teachers, challenged the system, arguing that the process was unconstitutional. By June 2013, the state deemed that using student growth data in this manner was no longer permitted as per Florida Senate Bill 980, passed in July 2013. However, a revised federal lawsuit was launched the following October, based upon allegations that some school districts were still using FCAT scores to evaluate teachers in other subject areas. A federal district court dismissed this lawsuit in May 2014. The judge ruled that, while the practice of evaluating teachers based upon students they did not have was not fair, it was not unconstitutional.

Table 1.01 summarizes the status teacher evaluation for each of the four states that are examined in this study (National Council on Teacher Quality, 2014). Beyond providing information about the use of achievement data used in evaluations, the table also summarizes implementation timelines, the role of classroom observation in teacher evaluations, and specific policy elements related to tenure/dismissal.

**Table 1.01. Teacher Evaluation Status Summaries for Target States.**

	<b>Tennessee</b>	<b>North Carolina</b>	<b>Rhode Island</b>	<b>Florida</b>
<b>Race to the Top Grant award date</b>	March 29, 2010	August 24, 2010	August 24, 2010	August 24, 2010
<b>First full school year of teacher evaluation implementation</b>	2011–2012	2012–2013	2012–2013	2011–2012
<b>Evaluation system structure</b>	Tennessee Educator Acceleration Model; Presumptive state evaluation model for districts with possible opt-out	North Carolina Educator Effectiveness System; Presumptive evaluation model for districts with possible opt-out	Teacher Evaluation and Support System; Presumptive evaluation model for districts with possible opt-out	State criteria or framework for district-designated evaluation system
<b>Use of achievement data/student growth in teacher evaluations</b>	State requires 50% of evaluations must be based upon student achievement, of which 35% must rely upon student growth from Tennessee Value-Added Assessment System	All teachers must be evaluated on six standards; A teacher cannot be rated effective if he/she does not meet expected student growth standard; Teacher effectiveness status determined over 3-year average	State student learning component accounts for majority of teacher evaluation score, based upon progress toward academic goals learning standards, and results from Rhode Island Growth Model	At least 50% of teacher evaluations must be based on data and indicators of student learning growth as measured by statewide assessments over the course of at least three years
<b>Observations</b>	Classroom observations required (minimum unspecified)	Classroom observations required (minimum unspecified)	Minimum of 3 per year	Required; New teachers must be observed at least twice in first year
<b>Tenure policy</b>	Tenure awarded to teachers whose overall effectiveness rating is above expectations or higher for two years; Tenured teachers with two consecutive ratings below expectations may be reverted to probationary status until they receive two consecutive ratings of above expectations or higher	Teachers employed less than three years are given one-year contracts; Contracts only offered to teachers demonstrating effectiveness/ proficiency on the evaluation instrument	Virtually impossible for ineffective teachers to achieve non-probationary status	To be awarded an annual contract, a probationary teacher must not receive: two consecutive annual ratings of unsatisfactory, two annual ratings of unsatisfactory within a three-year period, or three consecutive needs improvement or a combination of needs improvement and unsatisfactory ratings
<b>Dismissal policy</b>	Inefficient teachers, who have evaluations demonstrating overall performance that is below expectation, are subject to dismissal	Non-probationary teachers may be terminated for inadequate performance (failure to meet proficiency on any standard of evaluative instrument)	Districts required to dismiss all educators rated ineffective for two consecutive years	An annual contract may not be awarded if the teacher has received any of the ratings outlined in the tenure policy
<b>Considerations for music educators</b>	Fine Arts Growth Measure portfolio system for growth (not yet state wide)	Analysis of Student Work system for student growth (not yet state wide)	Music educators can be subject to RIGM if considered “contributing educator”	Performing Fine Arts Assessment Project for student growth (not yet state wide)

*Note.* Information in this table was drawn from the National Council on Teacher Quality policy briefs on state-by-state evaluation timelines, published online April 7, 2014 (National Council on Teacher Quality, 2014).

## Need for the Study

According to Schmidt and Datnow (2005), “Teachers are considered by most policymakers and school change experts to be the centerpiece of educational change” (p. 949). In particular, teacher buy-in is a critical element of successful reform implementation (Spillane, Reiser, & Reimer, 2002). Researchers have suggested that education reforms have stronger staying power when teachers feel connected to the process (Fullan, 1993; Schmidt & Datnow, 2005). In contrast, teachers may be resistant to reforms if they perceive such change threatens their values or interests (Muncey & McQuillan, 1996).

Though RttT guidelines allow for a vetted approach to teacher evaluation (e.g., pilot-testing approaches, tiered enactment of new policies in stages), educators must still adapt to new, rigorous standards and classroom practice expectations that may impact their professional identity, agency, and autonomy. The causal-based model of action underlying new teacher evaluation initiatives promotes a punitive accountability orientation that may be at odds with teachers’ motivational interests and professional development needs. The scope and rate of change posed by teacher evaluation reforms may also challenge teachers’ long-held beliefs and classroom routines, thereby contributing to a sense of professional vulnerability (Lasky, 2005). Long-term, teachers’ beliefs about and overall commitment to the profession may be influenced by RttT policies.

Historically, music educators have desired a fair, clear, and useful teacher evaluation process (e.g., Brophy, 1995; Shaw, 2013; Taebel, 1990a). By examining music educators’ perspectives on recently implemented evaluation systems, researchers may ascertain whether such systems do or do not align with music educators’ professional priorities, and the degree of music teacher buy-in that may exist. Currently, there is a dearth of research investigating

educators' perceptions of RttT-driven teacher evaluation systems (Hasty, 2013; Thomson, 2013), and while some researchers have examined music teachers' experiences with evaluation (Maranzano, 2000; Shaw, 2013; Taebel, 1990a), no researchers have specifically explored the impact of established, RttT-driven teacher evaluation policies from the perspective of K–12 music educators.

### **Study Purpose and Research Questions**

The purpose of this exploratory study was to examine music teachers' perspectives of current teacher evaluation practices designed to comply with RttT initiatives being implemented in four states. To date, this is the first multi-state survey of music educators and their perceptions of RttT teacher evaluation policies and processes. In this study, I addressed the following questions:

1. *What elements characterize the process used to evaluate music educators? By what measures are music teachers evaluated? Who is primarily in charge of conducting music teacher evaluations? In what ways are student achievement data considered when determining music teacher quality? What outcomes are associated with evaluation results?*
2. *How do music educators perceive the teacher evaluation process? Do music educators perceive teacher evaluation systems to be transparent and fair? Are evaluations seen as positive or punitive in nature? To what extent are music educators satisfied with the evaluation process by which they are held accountable?*
3. *In what ways do music teachers believe evaluation practices impact and/or benefit the classroom?*

Do music teachers believe teacher evaluation systems contribute to enhanced student

learning? Are evaluation procedures contributing to improvements in teaching practice? Do evaluation expectations influence classroom routines?

4. *How do teacher evaluation practices impact music educators' perceptions of music teaching as a profession?* Do music teachers believe teacher evaluation elevates or diminishes the status of music education?

### **Definitions**

1. Teacher evaluation: The overall process by which teacher competence and effectiveness is determined.
2. Race to the Top: Federal initiative encouraging states to redesign educational practices to better address elements such as effective teacher evaluation, struggling schools, and Common Core standards in exchange for federal funding earmarked for enacting proposed reforms.
3. Teacher evaluation objectives/purpose: The outcomes or functions associated with assessing teacher effectiveness. These functions can serve summative purposes (e.g., to facilitate personnel decisions) or formative purposes (e.g., to improve professional practice).
4. Measures of teacher performance: The sources of evidence used to determine teacher competence and quality in teacher evaluation (e.g., classroom observations, value-added data, student input/feedback).
5. Student growth: Records of student achievement across two points in time, such that student academic development can be determined.

6. Non-tested subjects: Content areas which do not have a systematic, statewide, standardized assessment designed specifically for them (e.g., physical education, music, consumer sciences) (Goe & Holdheide, 2011).
7. Primary evaluator: The person chiefly responsible for determining a teacher's overall quality during the evaluation process.
8. Evaluator credibility: The degree to which the primary evaluator is sufficiently knowledgeable and adequately trained to evaluate a given teacher's overall quality during the evaluation process.

### **Delimitations**

Study participation included K–12 music educators teaching in Tennessee, Florida, North Carolina, and Rhode Island. The study was delimited to practicing music educators in states who have received RttT grant funding in exchange for redesigning teacher evaluation practices. This delimitation was applied to examine the various iterations of teacher evaluation that complied with the federally funded grant reform stipulations. These four states in particular each implemented RttT-approved teacher evaluation systems for one year or more, and as such, reformed teacher evaluation practices in these states were more firmly established than practices in those states who earned RttT funding in later rounds. Therefore, music teachers in these states were able to provide more meaningful responses regarding their lived experiences in the face of new evaluation practices.

Participation was further delimited to those K–12 music teachers who were current members of NAFME. This organization is one of the largest associations affiliated with music education, and it operates at local, state, and national levels. Membership in the NAFME organization presumes a certain level of professional commitment and awareness of major

education policy initiatives. Therefore, delimiting the sample to NAfME members potentially contributed to more information-rich data about the nature music educators' perceptions regarding current teacher evaluation practices.

## Chapter II

### Review of Related Literature

*“Few issues in education are more explosive than evaluating teachers and teaching.”*

*(Finkel, 1983, p. v)*

#### Introduction

Prior to the 21<sup>st</sup> century, teacher evaluations emphasized subjective conceptualizations of “good” teaching, often characterized by ambiguous criteria and feedback that failed to foster sufficient professional development and growth. Particularly before the 1950s, evaluations typically depended upon high-inference classroom observation measures, yielding inconsistent results (Lavigne & Good, 2014). Evaluators were often seen as incompetent, practices and implementation were widely varied, and as a result, teacher evaluations were perceived as unreliable (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). Concerns regarding how teacher effectiveness was determined, coupled with the public’s growing demand for educational accountability, brought teacher evaluation into focus as a major reform initiative in recent years (Weisberg, Sexton, Mulhurn, & Keeling, 2009).

Policies enacted by Race to the Top’s (RttT) “Great Teachers and Leaders” frame an evaluative model that emphasizes quantifiable, objective evidence of teacher effectiveness. One example of such quantifiable evidence is student achievement data. As per RttT policy, student achievement data must constitute a significant portion of a teacher’s overall evaluation, a novel practice compared with past evaluation systems. The “Great Teachers and Leaders” provision also calls for prescribed, validated classroom observation instruments to be used by trained observers, in order to strengthen the reliability of classroom-based evidence.

Whereas RttT guidelines call for multiple measures of teacher effectiveness – beyond

student achievement data – to capture a comprehensive, nuanced picture of teacher quality, critics note, “the multiplicity of [these] measures represents merely a many-voices-in-unison expression of the same measure,” and further widens “the gulf between the quantitative estimates and the real-world sense of great teaching to which they are meant to speak” (Gottlieb, 2013, p. 19). In other words, critics are concerned about the validity and reliability of the measures used to practically ascertain teacher effectiveness, as well as the impact of those measures on actual teacher practice and development. To better understand the nature and utility of RttT’s systematic approach to teacher evaluation, it is critical to examine practicing educators’ perspectives on their lived experience under the umbrella of RttT teacher evaluation policy. Investigating teachers’ perspectives is particularly important given high-stakes sanctions associated with teacher evaluations and the impact such outcomes may have on teachers’ career satisfaction, motivation, and commitment.

During the past decade, several researchers have examined educators’ perceptions of the teacher evaluation process (e.g., Delvaux et al., 2013; Gratton, 2004; Heneman & Milanowski, 2003; Kelly et al., 2007; Milanowski & Heneman, 2001; O’Pry & Schumacher, 2012; Schumacher, 2010). In the wake of newer teacher evaluation legislation, however, very few researchers have investigated teachers’ perceptions specific to RttT teacher evaluation policies (Hasty, 2013; Thomson, 2014). Examining teachers’ perceptions of such policies is particularly critical, as RttT teacher evaluation requirements have strong implications on both procedural and distributive justice for practicing educators. Moreover, as RttT funding has been distributed over three phases and related teacher evaluation policy has been enacted by several states (18 to date), and given that the Obama administration’s position on teacher evaluation remains largely unchanged, it is possible that more states will continue to adopt teacher evaluation systems

modeled after those developed within RttT states. Beyond the dearth of research specific to evolving teacher evaluation approaches, no known studies examine the unique experiences of music teachers subject to RttT-driven evaluation policies. As these evaluation policies were largely designed with the general classroom teacher in mind, data specific to RttT's impact upon music teachers as specialists may reveal important or unique outcomes specific to the evaluative process, music educators' experiences with and perceptions of the process, teaching behaviors, and more global views of teaching as a profession.

The research reviewed in this chapter is organized into two main sections. The first section focuses upon studies examining teachers' perceptions of the overall evaluation process in general education. Within this section, I summarize research specific to teachers' perceptions of (a) the function of teacher evaluation, (b) measures of teacher effectiveness used in evaluation, (c) evaluator credibility and competence, and (d) the overall fairness, clarity, and utility of the evaluation process. I also summarize recent research specific to RttT-driven teacher evaluation policies, focusing on teachers' experiences with and reactions to new evaluation initiatives. In the second section, I explore research specific to music educator perceptions of the teacher evaluation process, with attention to studies highlighting the unique competencies of music educators, as well as studies focusing upon the extent to which those individuals evaluating music teachers have been properly trained.

### **Perceptions of Teacher Evaluation in General Education**

Interest in teacher perceptions of the evaluative process is rooted in research emphasizing the importance of employees' perceptions of their performance evaluation system (Keeping & Levy, 2000). In particular, those employees that find their evaluation process to be fair and just are more likely to find their evaluation satisfying (Folger, 1987), and such satisfaction may

predict overall organizational commitment (Konovsky, Folger, & Cropanzano, 1987). Research on teachers' perceptions of evaluation processes have focused upon evaluative functions (Gratton, 2004; Kelly et al., 2007; Nolin, Rowand, & Farris, 1994), measures of teacher performance (Attinello, Lare, & Waters, 2006; Danielson & McGreal, 2000; Nolin, Rowand, & Farris, 1994), and evaluator credibility and competence (Delvaux et al., 2013; Nolin et al., 1994; Zimmerman & Deckert-Pelton, 2003). Only limited research has focused upon teachers' perceptions of evaluative practices in the current RttT era (Hasty, 2013; Thomson, 2013).

**Teacher evaluation functions.** Evaluation systems typically serve two distinct functions. Evaluations can be used to help improve teachers' individual professional practice (i.e., evaluation for professional development, or *formative* evaluations), and teacher evaluation data can facilitate personnel decisions (i.e., evaluation for educational accountability, or *summative* evaluations) (Danielson & McGreal, 2000; Delvaux et al., 2013; Duke, 1990; Marzano, 2012; McGreal, 1983; Stiggins, 1986; Stronge, 2006; Wise et al., 1984). According to several researchers, *formative and summative* evaluation dimensions are critical to a comprehensive, effective teacher evaluation system (Colby, Bradshaw, & Joyner, 2002; Marzano, 2012; Stronge, 2006; Stronge, 1995). However, critics argue that these functions are inherently contradictory (Firestone, 2014). Although evaluation systems adhering to RttT guidelines purport to serve both of these functions, the renewed emphasis on teacher accountability has, in fact, introduced a *stronger* focus on *summative* functions that link evaluative outcomes with some combination of incentives and sanctions.

Gratton (2004) investigated a teacher appraisal process implemented in a large, urban school in New Zealand. A stratified random sample of 17 faculty completed a questionnaire examining teachers' perceptions of the evaluation process, including what they believed *should*

be the purpose of such evaluations. Seven of these faculty participated in additional follow-up interviews. In terms of evaluation function, participants indicated they believed the evaluation process leaned slightly more towards accountability purposes, which several teachers perceived as threatening. When responding to the prompt regarding what *should be* the primary focus of the appraisal process, participants overwhelmingly indicated that professional development ought to be the main purpose.

In examining informational documents regarding the evaluation process that had been provided to the building faculty, Gratton determined that teachers received only limited guidance regarding evaluation expectations. Questionnaire responses confirmed this finding. Participants were asked how clearly the purpose of the appraisal process had been explained to them, responding on a 5-point Likert-type scale ranging from 1 (*not clearly*) to 5 (*very clearly*). The mean score of 2.51 suggested that some teachers were somewhat confused about the evaluation process. Overall, teachers felt evaluations to be time consuming with few favorable returns, thereby negatively impacting their commitment to the process.

In an investigation of 85 primary school teachers in Singapore, Kelly and colleagues (2007) asked educators to indicate how important they believed nine possible evaluation functions were in the context of their performance appraisal system. A principal components factor analysis determined that three factors accounted for 68% of the variance among evaluation function items. Factor 1 included three items loading onto *formative purposes* (e.g., “to identify the professional development needs of teachers”), Factor 2 included four items loading onto *summative purposes* (e.g., “to hold teachers accountable for their performance”), and Factor 3 included two items loading onto *human relational purposes* (e.g., “to promote good working relationships among teachers”). Using a 5-point, Likert type scale ranging from 1 (*not important*

*at all*) to 5 (*extremely important*), participants gave the highest average importance rating to human relational purposes ( $M = 4.30$ ,  $SD = 0.73$ ,  $\alpha = .53$ ), followed by formative purposes ( $M = 4.27$ ,  $SD = 0.54$ ,  $\alpha = .79$ ) and summative purposes ( $M = 3.87$ ,  $SD = 0.70$ ,  $\alpha = .77$ ). Human relational and formative purposes were perceived as significantly ( $p < .001$ ) more important than summative purposes, in terms of overall evaluation goals.

In a more recent study, Flores (2012) surveyed 150 teachers in Portugal to determine educators' perceptions of a new evaluation process. Using an instrumental case method approach, Flores gathered both survey and interview data from teachers in one elementary and one secondary school in northern Portugal. Similar to findings by Gratton (2004), Flores determined that participants valued formative objectives over summative goals with regard to the evaluation process. The majority (67%) disagreed that evaluations should be used for managerial decisions. Interestingly, the majority of participants (54%) did not believe evaluations should be used for both professional development *and* accountability purposes, suggesting that teachers may sense the conflicts that may accompany dual-purpose evaluation systems.

Staff at the National Center for Education Statistics (NCES) administered a survey to approximately 1,000 United States elementary public school teachers to determine their perspectives on teacher evaluation (Nolin et al., 1994). This instrument, the "Survey on Teacher Performance Evaluation" (STPE), solicited participants' beliefs about the various functions associated with their evaluation process. For example, participants rated the extent to which they believed their evaluation system was used "to make tenure and promotion decisions" (using a 4-point Likert-type scale ranging from *not at all* to *a great extent*). Results indicated that 61% of participants believed teacher evaluations, to a great extent, served the objective of "guiding

improvement of teaching skills.” Only 18% of participants believed that the objective “to discharge incompetent teachers” was considered to a great extent.

Participants also indicated the extent to which they believed certain objectives *should be* considered in the teacher evaluation process. Results showed 45% of participants believed evaluations *should be* used to discharge incompetent teachers, while 81% believed evaluations *should be* used to guide the improvement of teaching skills. These results reflect the complex nature of evaluation systems designed to address both formative and summative objectives.

**Measures of teacher performance.** According to Danielson & McGreal (2000), “the cornerstone of any evaluation system is the set of evaluative criteria on which a school or district bases its teacher evaluations” (p. 32). Criteria serve as the standards by which excellent teaching is judged. To make determinations about high-quality teaching, data must be collected from a variety of sources, or *measures*, of teacher performance. These measures typically encompass either inputs (i.e., the tasks teachers engage in on a day-to-day basis) or outputs (i.e., the results of teachers’ efforts) (Danielson & McGreal, 2000). This distinction is important, as teacher evaluation processes may focus on the quality of the *teacher* and/or the quality of the *teaching* (Darling-Hammond, Wise, & Pease, 1983). Examples of input measures used in teacher performance evaluations include depth of professional knowledge and quality of instructional planning, while outputs are typically represented by student achievement (Danielson, 2014; Darling-Hammond et al., 1983; Stronge, 2010).

Drawing evaluative data from multiple measures increases the reliability and validity of a given evaluation system, and purposeful evaluation systems typically include measures of both inputs and outputs (Goldrick, 2002). A system based upon prescribed measures of effectiveness that are used to inform evaluative criteria is known as standards-based evaluation. Standards-

based evaluation systems facilitate comparisons across teachers who possess diverse forms of expertise or who work within distinct disciplines and contexts, but such systems may lack the nuance and flexibility necessary to yield valid and meaningful evaluation for all teachers.

In the “Survey on Teacher Performance Evaluation” (STPE) administered by the NCES, teachers were asked to report on the measures of teacher performance considered in their most recent evaluation (Nolin et al., 1994). For example, participants indicated the extent to which factors such as “instructional techniques” and “subject matter knowledge” were considered in their performance evaluation, using a 3-point Likert-type scale ranging from *small extent or not at all* to *a great extent*. The majority of participants (84%) believed “instructional techniques” were considered to a great extent in their most recent evaluation, while only 65% felt “subject matter knowledge” was considered to a great extent. Only 53% of participants believed “teaching demands unique to the students in the classroom” were strongly considered.

In the NCES study, participants also indicated whether or not they believed those same components *should be* considered in their next evaluation cycle. Results indicated that 99% of participants believed overall teacher performance, subject matter knowledge, classroom management, and instructional techniques *should be* considered to a great extent in the evaluation. The greatest disparity between what *was* considered and what *should be* considered surfaced with “teaching demands unique to the students in the classroom,” as 95% of participants believed this component *should be* considered to a great extent.

Other researchers have examined teacher perceptions regarding specific measures of teaching effectiveness. For example, Attinello et al. (2006) explored teachers’ and administrators’ beliefs regarding the usefulness of a portfolio-based measure of teacher evaluation. A total of 752 teachers and 46 administrators responded to a questionnaire regarding

the accuracy, utility, and comprehensiveness of portfolios used in teacher evaluations. Follow-up interviews and focus group discussions provided additional qualitative data regarding teacher perceptions on portfolio-based measure of effectiveness. Generally, administrators expressed more favorable impressions about portfolios than teachers. Both administrators and teachers believed portfolios more comprehensively represented teacher performance when compared with observation alone, though teachers expressed concerns about time requirements associated with portfolio development. Although quantitative results suggested administrators and teachers believed portfolios accurately reflected work in the classroom, qualitative findings revealed that both groups did, in fact, have some concerns regarding portfolio accuracy.

In a more recent study, Campbell (2013) investigated teachers' and administrators' perspectives on "mini" classroom observations within the context of teacher evaluation. Unlike traditional conceptions of classroom observation that emphasize one-shot measures of complete class periods, mini observations require the evaluator to observe the teacher on several occasions for brief periods of time. Campbell sought to determine whether teachers and administrators believed mini observations enhanced instructional practice and improved professional relationships. A total of 33 participants completed an anonymous, 19-item online questionnaire regarding mini versus traditional observations. Campbell also collected data from in-depth individual interviews with three teachers and three administrators, as well as through three focus group sessions (both unstructured and semi-structured), which were conducted throughout the school year. Findings suggested more frequent classroom observations enhanced the relationships between teachers and administrators. Furthermore, mini observations contributed to reduced stress and eliminated notions of the "dog and pony show" phenomenon, wherein teachers felt compelled to put on an act for administrators during less frequent, more high-stakes

observations. Teachers also appreciated the more immediate feedback associated with mini observations.

***Student growth measures.*** Among RttT states, student growth represents an additional measure by which teacher effectiveness is gauged. In teacher evaluation parlance, student growth refers to the degree to which a student improves in a given area of learning across two points in time. RttT policies require that teacher evaluations be *significantly* informed by student growth. To attribute student growth to an individual teacher, a variety of growth models may be considered. Some examples of growth models include: (a) growth-to-proficiency, in which initial student performance informs yearly growth targets benchmarked toward achieving proficiency in a given area, (b) projection, which compares projected achievements against target proficiency standards, and (c) value-added, through which student growth is assessed over two points in time, controlling for past achievements and background characteristics (O'Malley, Murphy, McClarty, Murphy, & McBride, 2011). Student growth is often considered part of teacher evaluations, even in states that did not earn RttT grant funding. In fact, as of 2013, a total of 35 states require that student growth be a significant factor in teacher evaluations (Doherty & Jacobs, 2013).

There is limited research on teachers' perceptions of the use of student growth data in their evaluations. In one study, the Tennessee Department of Education released a report on the first year of implementation of their new evaluation system after surveying teachers and administrators throughout the state (Tennessee Department of Education, 2012). The report revealed that "teachers in subjects and grades that [did] not yield an individual value-added score indicated that having 35 percent of their score based on school-wide data [was] not reflective of their performance," suggesting this approach is not widely supported by classroom teachers.

Around 36% of teachers received an individual value-added score during the 2011–2012 school year, while all other teachers' performance was determined by composite measures (i.e., building-wide student performance data). Delaware released a similar report on their first year implementation of RttT teacher evaluation policies (2010–2011). Summarizing data from local education agencies, Delaware policymakers reported comparable challenges in developing and implementing assessments for students in traditionally non-tested subjects, complicating the process of evaluating teachers in those areas.

**Evaluator credibility and competence.** Duke and Stiggins (1986) note that the credibility of individuals primarily responsible for observing and evaluating educators is critical to the validity and utility of the evaluative process. The authors point out that credibility is a function of several factors, including “recent teaching experience,” “knowledge of subject area,” and “familiarity with the teacher’s classroom and students” (p. 22). Those evaluators with elevated credibility are seen as more competent overall, thus better entrusted by teachers to conduct the evaluation process.

Many educators have expressed concern regarding the credibility and competence of those responsible for conducting teacher evaluations. For example, in the STPE administered by the NCES, educators were asked to indicate how confident they were that the individual responsible for their most recent performance evaluation could competently assess facets of teaching (Nolin et al., 1994). A majority of participants felt their evaluators were qualified to assess their overall teacher performance (73%) and their classroom management skills (74%), while fewer participants were confident in their evaluators’ ability to assess their subject matter knowledge (62%) and instructional techniques (68%). These findings suggests that teachers are more confident in their evaluator’s ability to assess general teaching competencies rather than

content-specific teaching behaviors.

Zimmerman and Deckert-Pelton (2003) examined teachers' perceptions of their administrator's role in evaluation practices. A total of 86 teachers completed the Professional Appraisal System Survey, an 11-item open-ended instrument designed to gather information regarding teachers' perceptions of effective evaluation system, how evaluation might be improved, and whether or not evaluations were executed in a consistent manner. A constant comparative analysis allowed for organizing participant responses by theme into general categories. All responses loaded onto four domains: (a) interaction, (b) consistency, (c) commitment, and (d) knowledge. Results indicated that participants craved more interaction and meaningful suggestions from their evaluator, supporting findings reported by Campbell (2013). Though many participants felt the evaluation process was consistent, some cautioned that lack of consistency could weaken teacher motivation. Participants also believed that the evaluator's commitment to the evaluation process enhanced its meaningfulness. However, participants had mixed reactions regarding their administrator's competence and knowledge as an educator. Although some teachers felt that administrators were a viable source for pedagogical information, most perceived administrators as lacking the pedagogical experience and/or content knowledge expertise necessary to be competent evaluators.

Kelly and colleagues (2007) (study described earlier) conducted a stepwise, multiple regression analysis to determine which elements predicted Singapore teachers' cooperativeness with the evaluation process. Ultimately, the evaluator's credibility, along with the evaluator-evaluatee relationship, were the strongest predictors of teachers' overall cooperativeness. In other words, those teachers who trusted their evaluator's competence demonstrated higher levels of cooperation.

In their 2013 study, Delvaux and colleagues administered a questionnaire to 1,983 teachers across 65 schools in Belgium to investigate various features of the teacher evaluation system, including leadership characteristics of the evaluator and the evaluator's overall credibility. Participants indicated the extent to which they agreed with statements such as, "My evaluator was familiar with the subject matter that I taught," using a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Overall, participants felt slightly more positive about their relationship with their evaluator ( $M = 4.17, SD = 0.92$ ) than they did about their evaluator's credibility ( $M = 3.75, SD = 1.01$ ). However, participants also felt most strongly about their relationship with their evaluator when compared with all other features of the evaluation process (e.g., clarity, utility). Teachers experienced greater effects on professional development when their primary evaluator was perceived as credible, finding their feedback more useful. When the primary evaluator had a positive attitude toward the evaluation system, teachers also benefitted more from the evaluative process.

**Fairness, clarity, and utility.** In addition to gathering data regarding teacher perceptions about the objectives, criteria, and evaluators associated with the evaluation process, many researchers have examined teacher beliefs about the fairness, clarity and utility of the performance evaluation process itself. For the purposes of this study, *fairness* refers to the just, valid, and applicable nature of the evaluation process. *Clarity* references how clearly various aspects of the evaluation process are made to those being evaluated, and *utility* refers to the evaluation process's usefulness and effectiveness in improving teaching. For example, several items on the STPE (administered by the NCES) addressed participants' perceptions regarding the fairness, clarity, and utility of their most recent evaluation. With regard to fairness, 89% of teachers felt their evaluation provided an accurate depiction of their performance. In terms of

clarity, 94% of participants felt the evaluative criteria was made clear to them before the evaluation process began. Similarly 74% of participants believed the evaluation process was useful in improving their teaching skills.

In a 2001 study, Milanowski and Heneman investigated teachers' perceptions of a standards-based teacher evaluation system as framed by the Danielson model.<sup>3</sup> A total of 291 teachers from 10 schools in a Midwestern school district completed a 23-item questionnaire, responding to items such as, "The evaluation I received was accurate," and "Overall, I have a good understanding of the evaluation system," using a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). In addition to the survey data, a total of 35 interviews were conducted with teachers who were completing a more comprehensive evaluation during that school year. An additional 23 teachers undergoing traditional annual evaluations were also interviewed. These interviews solicited participants' beliefs about the evaluator, as well as their perceptions of the clarity and fairness of the evaluation process.

Survey results suggested participants demonstrated generally positive beliefs about the clarity of the evaluation process ( $M = 3.89, SD = 0.93$ ) and the fairness of evaluation results ( $M = 3.80, SD = 0.96$ ). Participants were also relatively satisfied with the ways in which the evaluation process impacted their teaching practice ( $M = 3.21, SD = 0.96$ ). Interview results indicated that those undergoing comprehensive examinations were more concerned with issues of fairness and evaluator expertise than those undergoing lower-stakes annual examination.

Those participants involved in comprehensive evaluations were also comparatively more

---

<sup>3</sup> Charlotte Danielson, an American economist and specialist in teacher effectiveness, developed the *Framework for Teaching* (FFT) as a model to define good teaching. The model, originally conceived in 1996, was intended to encourage reflective practice and professional development among practicing teachers. The model is comprised of 22 components, aligned with the Interstate Teacher Assessment and Support Consortium standards, which represent four primary domains of teaching responsibility. In recent years, the model has also been implemented in teacher evaluations. Research has shown the model has some degree of predictive validity; that is to say, those teachers that perform well as per the FFT may have more successful students (Danielson, 2014).

confused by the evaluative process. Teachers across both the comprehensive and annual examination groups viewed the evaluative process as being less applicable to teacher performance in subject areas such as music and art.

Kimball (2002) conducted a qualitative case study among three school districts implementing standards-based teacher evaluation systems, also informed by the Danielson framework. In order to determine the perceived fairness and utility of the evaluation procedures, Kimball conducted interviews with a purposive sample of teachers with different levels of experience, involved at different stages in the evaluation process. A total of 55 teachers and 6 teacher evaluators were interviewed using semi-structured protocols, and interviews were analyzed and coded for common themes. Results indicated that participants with less teaching experience found the evaluation process more useful. Overall, teachers spoke favorably of the system's fairness and clarity, noting an appreciation for the input they were able to provide their evaluators regarding context-specific considerations and overall goal setting.

In a follow-up study, Heneman and Milanowski (2003) focused on teacher reactions to a new teacher evaluation system during the second and third years of implementation. One objective of this study was to gather more information regarding teachers' reactions to the evaluation system during its full implementation. To address this objective, researchers developed an instrument (Teacher Reaction Scale, or TRS) including nine 3-4 item subscales measuring teachers reactions to the new evaluation system, addressing facets such as utility, accuracy, cost-benefit, procedural justice, and satisfaction. Participants rated their level of agreement with statements in each category on a 5-point, Likert-type scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). A total of 164 teachers involved in a standards-based evaluation system in the Cincinnati, Ohio public schools participated across the 2001 and 2002

school years. Results indicated that participants had fairly strong negative reactions to items addressing satisfaction with the evaluation process, stress caused by the system, and effort exceeding benefit. Generally, participants accepted the standards and rubrics associated with the evaluation process as sound, though teachers in areas such as art and music believed these rubrics required modifications to be applicable to their classrooms.

Investigating the utility of the evaluation process, Colby et al. (2002) compared the experiences of teachers in districts using a traditional, state-mandated evaluation system with the experiences of teachers in districts using a locally developed evaluation system. The authors were interested in determining which group perceived the teacher evaluation process as having a more significant positive impact upon various school goals, including student learning and overall school improvement. Quantitative data were gathered using an amended version of the Teacher Evaluation Profile (TEP) created by Stiggins and Duke (1988). The 4,092 study participants that completed the questionnaire hailed from 21 school districts, with each district embracing a slightly different approach to teacher evaluation: (a) districts using the state mandated teacher evaluation system ( $n = 15$ ), (b) districts using a locally developed teacher evaluation system ( $n = 3$ ), and (c) districts using a hybrid system ( $n = 3$ ). Qualitative data was also gathered in the form of interviews with the personnel director from each of the participating districts, as well as from focus groups conducted with teachers in districts using locally developed systems. Results indicated that locally developed systems were seen as having a stronger positive impact on overall school improvement, professional development, and student learning when compared with state-mandated systems, as locally conceived goals and objectives were better aligned with district priorities. Furthermore, locally developed evaluation processes were perceived as more comprehensive and rigorous.

In their 2007 study, Kelly et al. (study described earlier) examined various attributes of the evaluation system as related to teachers' attitudes. The survey instrument also solicited teachers' beliefs regarding the overall desirability of the evaluation system, as well as the purposes or outcomes associated with the evaluative process. Using a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), participants indicated the extent to which they agreed or disagreed with various items describing attributes of the evaluation process. A principal components analysis demonstrated that, among 21 items, five factors explained 71% of the total variance. These factors included (a) fairness and clarity of the appraisal system ( $\alpha = .91$ ), (b) controllability of the appraisal criteria ( $\alpha = .62$ ), (c) teacher participation in developing appraisal system ( $\alpha = .52$ ), (d) discussion of appraisal system with teachers (one item only), and (e) appraiser-appraisee relationship and appraiser credibility ( $\alpha = .89$ ). Overall, 89% of participants believed it was desirable to have an evaluation system in place. A stepwise multiple regression analysis revealed that *fairness and clarity*, as well as *controllability of the criteria*, were the strongest predictors of teachers' overall satisfaction with the evaluation system. These two factors were also the most important predictors of job satisfaction and professional motivation.

In their 2013 study (described earlier), Delvaux et al. also sought to uncover which features of teacher evaluation systems (e.g., perceived fairness, overall clarity) most affected Belgian educators' overall professional growth and improvement. Total scores for the utility variable, "effects on professional development," revealed general ambivalence regarding the impact of teacher evaluation upon professional development ( $M = 2.96$ ), suggesting that teachers did not find the process particularly useful. However, those teachers with fewer than five years of teaching experience, as well as those teachers who found the evaluation criteria and purposes

clear, reported that evaluations had a greater effect upon their professional development. In another international study, Flores (2012) (described earlier) found that Portuguese teachers felt threatened by the evaluation process, with over 80% of participants suggesting that evaluations were unfair and unreliable. Moreover, participants largely believed that the evaluation process did not effectively contribute to improved teaching. Qualitative interview data supported skepticism evident in the quantitative questionnaire results, with one participant contributing, “Teachers are unmotivated and discontent” (p. 360).

Some researchers have specifically investigated *early* career teachers’ beliefs about teacher evaluation. O’Pry and Schumacher (2012) investigated first-year teachers’ perceptions of a standards-based teacher evaluation system in metropolitan Houston, Texas school district. Among 137 new district teachers, a total of 121 responded to a 24-item electronic questionnaire soliciting participants’ opinions regarding the Professional Development Appraisal System (PDAS). Participants responded to each item using a 4-point Likert-type scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). First-year teachers thought favorably of the PDAS model, with the vast majority (92%) of participants either agreeing or strongly agreeing to being “generally supportive” of the process. A majority of participants (59%) also agreed that the PDAS process was stressful, despite 62% offering they would participate in the PDAS model even if it were not required of them.

O’Pry and Schumacher then created composite satisfaction scores by summing response totals from the initial questionnaire. From these composite scores, the researchers selected 12 respondents who demonstrated the most positive and most negative perceptions of the PDAS process and engaged in follow-up interviews to determine the factors contributing to those teachers’ opinions. Participants were asked questions such as, “What are your thoughts about the

PDAS observation that was conducted in your classroom this year?” Each participant interview was recorded, transcribed, and coded inductively. A set of six themes and 31 sub-themes were then used to analyze all interview responses. Those participants dissatisfied with the PDAS model cited issues with confusion surrounding the process, the process’s relevance, and the process’s utility. Conversely, those participants satisfied with the PDAS model felt the process was helpful in improving their teaching process.

**Educators’ perceptions of RttT evaluation policy.** Few researchers have examined teachers’ perceptions of the evaluation process as specific to RttT policy. The limited amount of research in this area is largely due to the recent implementation of RttT teacher evaluation policies. Moreover, only those states interested in obtaining RttT funding revised their teacher evaluation systems to adhere to RttT guidelines. Therefore, a relatively limited amount of practicing educators are subject to RttT policy mandates.

Thomson (2013) investigated Massachusetts teachers’ perceptions of RttT-driven teacher evaluation policies and procedures. A total of 52 teachers involved in pilot testing Massachusetts’ new teacher evaluation system completed a questionnaire. From this sample of participants, nine subjects participated in a small follow-up focus group. Thomson sought to determine how teachers perceived the new Massachusetts teacher evaluation system, in terms of its validity, usefulness, and applicability to improved teaching practice. Thomson also examined whether or not teachers believed new evaluation policies impacted their teaching practice. Using a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), participants indicated their level of agreement with statements such as, “The [evaluation] rubric is a comprehensive representation of effective teaching practice.” Follow-up focus group interviews allowed participants to contribute more detailed information to support their answers on the

written questionnaire. Ultimately, the majority of participants (75%) believed that their evaluators collected sufficient evidence to determine their overall evaluation. However, teachers felt relatively neutral in terms of evaluation policies impacting their classroom teaching practice ( $M = 2.92$ , range 1–6) and improving overall professional practice ( $M = 3.29$ , range 1–6). Follow-up focus group interviews revealed participants' concerns regarding the complexity of the evaluation system and the time demand of evaluation requirements, although participants appreciated the opportunity to provide their own evidence of student growth.

In her 2013 dissertation, Hasty examined how educators responded to new teacher evaluation procedures adopted in the RttT era. Hasty interviewed a total of six teachers and two principals in two different schools within the same Michigan school district. Although Michigan was not awarded federal RttT funding, teachers in the selected district were responsible for implementing the newly adopted Common Core Standards into their curriculum, and their teaching performance was evaluated using the Danielson framework. All participants had been evaluated using this new model at least once prior to Hasty's study. Interviews focused upon questions surrounding teachers' and principals' interpretation of and response to the new teacher evaluation process. Interview questions also centered on the role principal leadership had in addressing teachers' responses to new evaluation procedures. Findings highlighted the importance of trust between the teachers and their principal-evaluator. Participants reported mixed perspectives on the role of teacher evaluation supporting their instruction, and concerns regarding the consistency and objectivity of the evaluative process were also discussed. Finally, participants expressed concerns regarding the use of student growth data in overall evaluations.

**Section summary.** In general education, there has been extensive research on teacher evaluation policy, including studies that have revealed teachers' perceptions of the evaluation

process. In terms of evaluation function, teachers seem to largely value formative over summative goals (Flores, 2012; Gratton, 2004; Nolin et al., 1994). Teachers appreciate competent, consistent, and credible evaluators (Delvaux et al., 2013; Zimmerman & Deckert-Pelton, 2003), but teachers are not always confident in their evaluator's ability to judge content-specific knowledge or instructional strategies (Nolin et al., 1994). This lack of confidence may translate to concerns regarding the fairness, clarity, and utility of the evaluation process (Milanowski & Heneman, 2001). When teachers are confused about evaluation processes, they tend to respond less favorably (Flores, 2012; O'Pry & Schumacher, 2012). Fairness and clarity of the evaluation process significantly contribute to teachers' overall satisfaction with annual evaluations (Kelly et al., 2007). Overall, researchers have highlighted a range of criticisms and concerns with teacher evaluation, and these interpretations are based largely on self-report data (e.g., questionnaires, interviews, focus groups) obtained from teachers themselves. Teachers' perceptions of evaluation reform as characterized by Race to the Top guidelines suggest additional concerns with the complexity and consistency of the more robust and systematic evaluation process mandated by federal funding (Hasty, 2013; Thomson, 2013).

### **Evaluation and Professionalism**

Darling-Hammond (1990b) describes teacher professionalism as a student-oriented, knowledge-based endeavor based upon mutual conceptions and collective enforcement of professional standards. Shulman (1998) adds that professions require "the exercise of judgment under conditions of unavoidable uncertainty" and "the need for learning from experience as theory and practice interact" (p. 516). Recently, stakeholders have expressed concern regarding how regulatory structures such as teacher evaluation might impact teaching as a professional community. Wise and Darling-Hammond (1985) highlight the connection between teacher

evaluation systems and teacher professionalism, arguing that evaluative systems “are credible only where teachers are considered a professional resources rather than an object of bureaucratic scrutiny” (p. 28). The ritual of teacher evaluation is often seen as a hoop through which teachers and administrators must jump, rarely contributing to improved professional practice. Wise and Darling-Hammond suggest that evaluations must be credible, and associated rewards and sanctions must be valued by teachers in order for evaluations to effectively support professional growth. According to Wise, et al. (1984), “[Teacher evaluation systems] can either reinforce the idea of teaching as a profession or it can further deprofessionalize teaching, making it less able to attract and retain talented teachers” (p. v). Darling-Hammond (1990b) argues:

Indeed, the very definition of ‘professionalism’ in teaching has been turned on its head in public schools. Rather than connoting a high level of training and knowledge applied to a practice that must, above all else, serve the needs of clients in intellectually honest ways, the term is used by most policymakers and administrators to mean an unquestioning compliance with agency directives. Evaluation criteria stress good soldiership and conformity with district policies rather than knowledgeable advocacy of appropriate teacher practices. The ‘professional’ teacher in common parlance is one who ‘does things right’ rather than one who ‘does the right things.’ (p. 31)

Only a limited amount of research has explored the connection between teacher evaluations and feelings of professionalism. Del Savio (1992) conducted a study to determine the degree to which teacher evaluation systems aligned with the definition of teaching as a profession. Del Savio developed an index of professionalism relevant to teaching, indicating that

(a) teacher autonomy, (b) responsibility for student achievement, and (c) peer participation in maintaining standards of practice were each integral in determining whether or a teacher evaluation system reinforced teaching as a profession. Among Connecticut teachers, Del Savio determined that, while many teacher evaluation plans reflected processes that might bolster teachers' professional autonomy, certain activities found critical to reinforcing teachers' sense of professionalism, such as self-assessment, were not as prevalent as expected.

In a later study, Reult (1998) examined growth-oriented teacher evaluation and its impact upon teachers' sense of professionalism. Using a qualitative approach, Reult interviewed six teachers and three principals in two Washington school districts with contrasting organizational contexts (accountability-oriented versus professional autonomy-oriented). Results indicated that, when implemented in a context promoting professional autonomy, growth-oriented evaluations more effectively facilitated professional growth. However, when implemented in a more controlled and accountability-centered context, growth-oriented evaluations were seen mostly as a mechanism for satisfying a bureaucratic agenda. Supporting arguments made by others researchers (Darling-Hammond, 1990a; Wise & Darling-Hammond, 1985), these findings suggest the broader context in which evaluations are implemented bears strong implications on how well teachers receive new evaluation practices. The context may also impact how useful evaluations ultimately are when helping teachers grow and develop professionally.

**Section summary.** Teacher evaluation practices may have an impact upon educators' sense of professionalism. In particular, the bureaucracy associated with teacher evaluation systems may detract from teachers' sense of professional autonomy, potentially deflating feelings of professionalism. In order for teacher evaluation systems to effectively contribute to

teachers' professional growth and sense of commitment to the profession, policies must value teachers' specialized knowledge and understanding of both their content area and their students.

### **Music Educators' Perceptions of Teacher Evaluation**

Though extensive research has been conducted on teacher evaluation in general education, more limited research focuses specifically on music teachers' perceptions of the evaluation process. A few researchers focused specifically upon important music teacher competencies to consider during evaluations (Baker, 1981; Smith, 1985), while other researchers have examined music teachers' response and reactions to evaluation practices in different states (Maranzano, 2002; Shaw, 2013; Taebel, 1990a). While many music educators are addressing the challenges presented by applying various evaluation models to music classroom teaching (Emert, Sheehan, & Deitz, 2013; Gerrity, 2013; Monigold, 2013; Perrine, 2013; Stamm, 2013), a profession-wide response to these challenges has yet to take hold.

**Evaluative instrument and procedures.** In an early study, Baker (1981) examined the competencies evident in effective music teaching to create a music-specific checklist to be implemented in classroom observations. Baker administered an initial questionnaire to 119 music teachers and administrators in Oregon, who together offered 689 competencies reflective of effective music instruction. These competencies were subsequently organized into six broad categories: (a) instructional skills, (b), interest in work and pupils, (c) classroom management, (d) musical scholarship and musicianship, (e) personality qualities, and (f) quality of concert performance. Music experts then identified the 30 most critical competencies for use in vocal, general, and instrumental music classroom scenarios, creating three distinct rubrics. Baker determined that 16 competencies were common to all three areas of music instruction, with each area (e.g., vocal, general, instrumental) having several competencies exclusive to that music

discipline. These findings suggest that, even within music, there exists variation in what constitutes effective music teaching.

Smith (1985) also explored teaching competencies specific to music educators. In a survey of members of the Florida Music Educators Association, participants indicated whether or not they viewed each of 186 competencies as “necessary” or “unnecessary” for first-year music teachers. These competencies included, “To play a prepared solo on a guitar at a beginning level,” “To demonstrate the ability to read in each of the seven clefs,” and, “To demonstrate a knowledge of diction by singing in at least three of the following languages: English, Latin, German, Italian, French, Spanish.” Of competencies listed, nearly 96% received a necessity rating of 70% or higher. These findings suggest that music educators believe a widely extensive range of skills is necessary for effective music teaching.

Taebel (1990a) investigated music teachers’ perceptions of the teacher evaluation process implemented in Alabama, while also comparing music teachers’ classroom observation performance competency scores against those scores of non-music classroom teachers. Music teachers (N = 152) rated various teaching competencies with regard to their applicability to music teaching on a 4-point Likert-type scale ranging from 1 (not at all appropriate) to 4 (very appropriate). Among the competencies listed, participants believed that “communicating clearly” and “knowledge of subject matter” were most relevant to music teaching ( $M = 3.8$ ). Participants also self-rated their own proficiency in each area of teaching competency on a scale from 1 (*strongly disagree*) to 10 (*strongly agree*). These self-ratings were then compared against each participant’s principal ratings of that participant for the same competencies. Participants’ average self-ratings of their teaching competencies were higher across nine out of ten competency categories. Principals believed music teachers were slightly more competent in “knowledge of

subject matter.” Teacher participants suggested that festival performances and community activities should also factor into their overall evaluation. Additional findings from this study included classroom observation data that demonstrated music teachers’ competency scores were significantly lower than those scores of general classroom teachers across 7 out of 10 competencies, suggesting a need to further research both the measures/tools used to evaluate music teachers, as well as the credibility of those individuals performing music teacher evaluations.

Maranzano (2002) investigated music educators’ perceptions of teacher evaluation practices in Virginia by distributing a questionnaire to a stratified random sample of members from the Virginia Music Educators Association during the 2001–2002 academic year. Of the 200 music teachers invited to participate in the study, 138 (69%) eligible participants completed and returned the questionnaire. Participants were asked questions regarding the nature of the methods and observations included in teacher evaluations, as well as questions regarding the applicability of evaluation for music teachers. The majority of participants (57%) believed the building principal was primarily responsible for evaluating music teachers, and 100% of teachers reported direct formal or informal observation was part of the evaluative process. Only 2% of participants indicated a music-specific form was used for in-class observations. Participants believed their evaluation results were moderately accurate and contributed moderately to professional growth. Overall, participants expressed the most concern over their evaluator’s expertise, the evaluative criteria, and the consideration of musical teaching concepts and pedagogy in the evaluation process.

In a more recent study, Shaw (2013) conducted a study among K–12 band and orchestra

directors in order to gauge music educators' impressions of teacher evaluation practices in Michigan. A total of 330 participants responded to an online questionnaire soliciting teachers' perspectives on various aspects of the evaluation process. Participants responded to items such as, "Music teachers cannot be fairly evaluated using generic evaluation tools," and, "Administrators seem to understand what I am doing in my music classroom," using a 4-point Likert-type scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). Other items solicited participants' preferences regarding who should be involved in their overall evaluation (e.g., an experienced music teacher from outside the school district, a local college music education professor). Because Michigan legislation mandates value-added components be considered in teacher evaluations, the questionnaire also included items regarding participants' beliefs about viable options for measuring student growth in music. Additional items addressed procedural components of the evaluation process, such as who was primarily responsible for conducting evaluations and how frequently classroom observations occur throughout the school year. Findings suggested that participants largely believed that music teachers could not be fairly evaluated using generic evaluative instruments and that administrators were typically not qualified to evaluate music teachers. Participants noted a preference for in-person evaluations conducted by music specialists. Echoing findings by Colby et al. (2002), teachers advocated for a local, tailored approach to music teacher evaluation.

Drawing upon research studies conducted over an extended period of time, music education scholars and other professional educators have produced summaries on the status of teacher evaluation policy how such legislation impacts music educators in various states, including Florida (Perrine, 2013), Indiana (Gerrity, 2013), Minnesota (Orzolek, 2014), Ohio (Monigold, 2013; Stamm, 2013), and Pennsylvania (Emert et al., 2013). Music teachers have

responded to teacher evaluation policies by creating music-specific standardized assessments to measure student growth (e.g., Florida) as well as through developing clearer rubrics to assess music teacher classroom performance during in-person observations (e.g., Indiana). Leaders within state music education associations may be involved with formulating positions statements and adapting teacher evaluation policies in ways that enhance their applicability to music teaching (e.g., Parkes, 2011), though the National Association for Music Education (NAfME) has also made deliberate efforts to address music-specific needs in teacher evaluation.

For example, with the support of NAfME, Doerksen (2006) provided a specific set of guidelines for appropriately appraising music teachers in his book *Evaluating Teachers of Music Performance Groups*. He asserted that evaluator area of expertise should not be a factor in effectively assessing the classroom performance of music educators and proposed that, with the proper training, any educator could effectively evaluate music teachers. Doerksen emphasized that regular communication between the evaluator and evaluatee, proper goal setting, and appropriate evaluative consequences all brought more meaning to the evaluative process.

More recently, in 2013 NAfME crafted and released a teacher evaluation workbook designed around the Danielson framework. This workbook was created to help music teachers advocate for valid, reliable, and reasonable approaches to teacher evaluation policy implementation, and to assist principals and other non-musician administrators responsible for evaluating music teachers in conducting observations that adequately capture the unique characteristics of effective music teaching commonly evident in both general music and secondary rehearsal contexts.

**Evaluator training and credibility.** In addition to investigating the processes and

measures associated with evaluation, researchers have specifically examined the training credibility and expertise of those evaluating music instruction. Although some researchers have found that it may be important for evaluators to have music-specific expertise, (Duke & Blackman, 1991; Schmidt, 1992), others believe music expertise is not necessary to effectively evaluate music teachers (Goddard, 2004).

Duke and Blackman (1991) investigated the performance ratings given by 100 music majors and 100 non-music, education majors to a general music teacher featured in a 12-minute video segment. Using a 6-point scale, each participant rated the teacher's performance on the following teaching domains: (a) reinforces correct responses, (b) gives corrective academic feedback, or none needed, (c) reinforces appropriate behavior, and (d) gives corrective social feedback, uses techniques to stop inappropriate behavior, or none needed. Participants also provided an "overall performance" rating of the teacher using the same 6-point scale. Results revealed that non-music education majors' mean ratings were significantly higher than music majors' ratings for three of the four domain areas. There was not a significant difference between mean ratings for the domain, "gives corrective social feedback." These results are consistent with other research determining that the extent of observer-evaluator's training in the content area observed may impact the overall quality of teacher evaluation as well as the magnitude of ratings assigned.

Schmidt (1992) examined the reliability of untrained observers' ratings of applied music instruction. Even though the observers were musically trained, they had not been specifically trained in using the evaluative instrument implemented in the study. Schmidt employed an adapted version of Abeles' (1975) Applied Teaching Rating Scale (ATRS) as the evaluative measure in the study, supplementing the 30 validated items from the ATRS with several

researcher-developed items on the nature of teacher feedback. Schmidt investigated the test-retest reliability and inter-rater reliability of the ATRS as determined by the untrained observers' exposure to 25-minute samples of applied music teaching. Findings revealed that, while some evaluative items had satisfactorily reliable alpha coefficients ( $\alpha > .60$ ), more items had comparatively low alpha coefficients ( $\alpha < .60$ ). Schmidt proposed that rater characteristics (beyond music-specific expertise) might offer explanation for evaluative items with low inter-rater agreement and/or low test-retest reliability. The broad range in findings also suggest that, even when evaluators have music-specific expertise, it is critical that evaluators are properly trained in using evaluative instruments to ensure consistently reliable appraisals of teacher effectiveness.

Goddard (2004) investigated elementary music teachers' and principals' perceptions of music teacher evaluation in Canada. In addition to completing a survey about music teacher evaluation, 14 participants also engaged in interviews about their experiences with the evaluative process. Finally, participants each evaluated a music teacher via videotape. Survey and interview data suggested music teachers and their administrators agreed that music educators should be evaluated as specialists, taking into consideration the unique nature of the music classroom. However, both groups found it unnecessary for music teacher evaluators to be music professionals themselves. Still, when evaluating the videotaped music educator, the music teachers offered more content-specific suggestions for improvement.

Hirokawa (2013) compared music teacher evaluations as conducted by observers with varying levels of music expertise and evaluative training. In the study, Hirokawa aimed to answer two main research questions:

1. To what extent do evaluation scores and attitudes differ based on observer background in evaluation techniques and/or in music?
2. To what extent does training in the evaluation of a music teacher affect the attitude toward music teacher evaluation and/or the actual evaluation completed by an observer with little or no musical background? (p. 58)

To answer the first research question, Hirokawa solicited participants from Pennsylvania who were responsible for evaluating music teachers, including (a) administrators with no musical training or background, and (b) administrators with musical training or background. A third group, music supervisors/leaders, was also included in the analysis, although not all of these participants were responsible for evaluating other music teachers. A total of 63 participants were involved in the study. Each participant reviewed the Teacher Observation Rubric (TOR), then watched a 15-minute video of a music teacher before completing the TOR and a subsequent attitudinal/demographic survey. Using a series of one-way ANOVAs, Hirokawa determined that, across the seven domains outlining the TOR (e.g., classroom environment, assessment, planning), no significant difference was found between the scores of musically trained observers and non-musically trained observers. However, certain domains, such as “Content Knowledge” ( $p = .05$ ) and “Context Specific Characteristics” ( $p = .05$ ) approached significance, suggesting that musically trained observers may differ from non-musically trained observers in certain areas. Musically trained observers were also able to offer more content-specific feedback, which may suggest musically trained observers can more contribute to music educators’ professional development in more effective, relevant ways.

To answer the second research question, Hirokawa employed a one-group pretest-posttest design to determine if a 75-minute training session specific to observing music classroom teachers in the ensemble setting would impact observers' impressions of the music teacher's performance. The training focused upon music-specific examples of each of the seven domains outlined in the TOR, and a total of five observers participated in this phase of the study. After the experimental intervention, raw scores remained relatively stable for four out of five participants. The fifth participant's raw score increased dramatically. Scores in each of the domain areas increased from pretest to posttest. The training seemed to most impact scores in the areas of "Content Knowledge" and "Instructional Strategies," suggesting these may be areas of focus for future studies examining training interventions in the context of music teacher evaluation. However, because of the small sample size involved in this portion of the study, results should be interpreted with caution. Ultimately, Hirokawa's study suggests that while the evaluation scores of musically trained and non-musically trained observers may not differ dramatically from one another, musically trained observers may provide more useful feedback. Moreover, providing additional training for music teacher observers may increase the usefulness of the evaluative process for music educators.

**Section summary.** Ultimately, researchers have revealed an overarching concern for the validity and fairness of applying generic evaluation systems to the unique context of music teachers' professional practice (Doerksen, 2006; Maranzano, 2002; Shaw, 2013; Taebel, 1990b). This concern stems largely from the notion that music teacher competencies may differ from those of traditional classroom teachers (Baker, 1981; Smith, 1985). Moreover, there is mixed evidence regarding the importance of evaluators' music-specific expertise (Duke & Blackman, 1991; Goddard, 2004; Hirokawa, 2013; Schmidt, 1992). Researchers have demonstrated that

evaluating music teachers is a complex and nuanced process. Ultimately, the implementation of evaluation policy with music educators, as well as teachers of other non-tested subjects, may need to be approached with particular care. Special efforts and resources may be required to assist principals and other non-music administrators in evaluating music teachers that are reliable, valid, and useful.

## **Chapter III**

### **Methodology**

#### **Introduction**

In an effort to promote greater accountability among state and local education agencies nationwide, the federal Race to the Top (RttT) fund was established. This fund provided substantial grant money to states that revised various components of their education system. In particular, to qualify for RttT grant funding, states were required to amend their teacher evaluation protocols. These revised protocols were largely designed with the general classroom teacher in mind. As such, music teachers subject to RttT evaluation practices may face unique challenges in meeting new evaluation expectations. The purpose of this study was to explore how teacher evaluation policies designed to comply with RttT-funded initiatives were implemented at the classroom level. An additional purpose of this study was to investigate music educators' impressions of RttT-driven teacher evaluation policies.

#### **Sampling**

To better understand (a) how current teacher evaluation policies are implemented in the music classroom, (b) how music teachers perceive these policies in terms of clarity, fairness and utility, and (c) the degree to which music teachers may be adapting their professional decision making and practice in response to these policies, I surveyed K–12 music educators teaching in states that were awarded RttT funding. I initially delimited my target population to five states where RttT evaluation practices had been implemented for one year or more. These states included Florida, North Carolina, Rhode Island, Delaware, and Tennessee.<sup>4</sup> According to the United States Government Accountability Office (USGAO) (United States Government

---

<sup>4</sup> Due to population access issues, I was unable to collect data from music teachers in Delaware. Therefore, Delaware was not included in this study.

Accountability Office, 2013), each of these states had largely implemented their newly revised teacher evaluation systems (though, not necessarily the student growth component or the use of evaluations to inform personnel decisions) by the 2012–2013 school year. Other states (such as Georgia, Maryland, and Ohio) only partially implemented or piloted their new teacher evaluation systems by the 2012–2013 school year. Because the purpose of this study was to explore music teachers' experiences with *enacted* teacher evaluation policies, music educators in the remaining RttT states – plus the District of Columbia – were not surveyed.

Table 3.01 summarizes states earning RttT grants in the first two rounds, the award amounts, and each state's implementation timeline. The table illustrates three distinct phases of each state's implementation timeline. The first phase involved the enactment of the new teacher evaluation system in general. For example, Rhode Island implemented its Teacher Evaluation and Support System model, which included professional foundations (e.g., instructional planning) and professional practice (e.g., effective instruction in the classroom) as part of a larger model. The second phase of each implementation timeline included using student growth data to inform teacher evaluations. In Rhode Island, student growth was determined using Student Learning/Outcome Objectives (SLOs and SOOs) and the Rhode Island Growth Model (RIGM). The final aspect of implementation involved using teacher evaluation data to inform personnel decisions, including promotion and retention.

**Table 3.01***States Awarded Race to the Top Grant Funding in the First Two Rounds*

State	Award Date	Award Amount	Implementation of Teacher Evaluation System	Implementation Including Student Growth Component	First Year Ratings Must Inform Personnel Decisions
Delaware	March 29, 2010	\$100 million	2011–2012	2012–2013	2012–2013
Tennessee	March 29, 2010	\$500 million	2011–2012	2012–2013	2013–2014
District of Columbia	August 24, 2010	\$75 million	2013–2014	2014–2015	2014–2015
Florida	August 24, 2010	\$700 million	2011–2012	2011–2012	2012–2013
Georgia	August 24, 2010	\$400 million	2013–2014	2013–2014	2015–2016
Hawaii	August 24, 2010	\$75 million	2013–2014	2013–2014	2014–2015
Maryland	August 24, 2010	\$250 million	2013–2014	2013–2014	2016–2017
Massachusetts	August 24, 2010	\$250 million	2014–2015	2014–2015	2014–2015
New York	August 24, 2010	\$700 million	2012–2013*	2013–2104	2012–2013
North Carolina	August 24, 2010	\$400 million	2012–2013	2015–2016	2016–2017
Ohio	August 24, 2010	\$400 million	2013–2014	2013–2014	2014–2015
Rhode Island	August 24, 2010	\$75 million	2012–2013	2013–2014	2012–2013

*Note.* Data drawn from the National Council on Teacher Quality evaluation timeline briefs (National Council on Teacher Quality, 2014) published online April 7, 2014. Music educators in shaded states were included in the target population. Although Delaware was also a target state, population access issues prevented the inclusion of Delaware in this study. \*In New York, evaluation policy implementation was subject to collective bargaining at the district level, which complicated and delayed complete statewide implementation in 2012–2013.

In general, the majority of states enacted the basic structure of their new evaluation systems prior to incorporating student growth components and personnel sanctions. Delaying the student growth component allowed states more time to ensure measures of student growth were adequately valid and reliable, while delaying the use of evaluation scores for personnel decisions allowed states time to first accumulate the necessary student growth data for each teacher.

After identifying the states that had implemented new evaluation protocols for two years or more, I further delimited my target population to those practicing K–12 music educators who were also active members of the National Association for Music Education (NAfME). With over 60,000 members, this organization is the largest professional association in music education, and at the state level, members are affiliated with state Music Educators Associations (MEAs). Membership in these MEAs presumes an elevated level of professional engagement and awareness of major education policy initiatives. Therefore, this purposeful, non-probability sampling strategy focused upon (a) states where RttT evaluation practices had been implemented for one year or more, and (b) K–12 music educators within those states that were also members of NAfME. To survey NAfME members, it was necessary to pursue a series of sampling access options.

**Data management company option.** In October 2013, I called the data management company responsible for handling all NAfME membership lists, in order to obtain access to the K–12 music teachers that constituted my target population. A representative from the data management company replied within 24 hours and provided me a preliminary quote of \$628.68, which would cover the cost of mailing labels. The representative also provided an order form, which required I submit a copy of the questionnaire I planned to send participants. The representative informed me that once the NAfME office approved my questionnaire, she would proceed with my order and provide me with the mailing labels.

The questionnaire was finalized in January 2014. After speaking with the same representative on the phone in late January, I completed my order form and submitted it, along with my finalized questionnaire, to the data management company on January 31<sup>st</sup>, 2014. The representative responded within 24 hours via email, and assured me I would hear from her within a week. On February 11<sup>th</sup>, 2014, I followed up with the data management company, and the representative informed me she was still waiting for the questionnaire to be approved. I followed up a second time the week of February 17<sup>th</sup>, 2014 via phone, left a message with the representative, but received no response.

**NAfME national office option.** Because I was concerned about my data collection window overlapping with music educators' spring breaks and the ensuing impact that overlap might have on my response rate, I pursued alternate routes to gaining access to my target population. During the week of February 17<sup>th</sup>, my primary advisor and I engaged in preliminary discussions with the NAfME President-Elect regarding a potential route to distributing my questionnaire directly through the national office. I composed a letter requesting NAfME's assistance, and I emailed the letter and a copy of the questionnaire to the NAfME national office. At the time, the executive committee was drafting a new policy that would help scholars conduct NAfME-approved research through the national office. However, after speaking with an executive at the NAfME office, I was informed that the committee's proposed timeline for this policy's implementation also conflicted with my data collection goals. As a result, I moved on to a third option, which involved contacting MEA representatives for each of my five target states, with the intent of distributing the questionnaire directly through the state MEA offices.

**State MEA option.** In order to distribute my questionnaire through the five individual state MEAs, I drafted an introductory email script explaining the nature of the study (See Appendix A), and I sent this script along with a .pdf version of the online questionnaire to the

MEA president and executive director in each of my five target states on February 25<sup>th</sup>, 2014. As a gesture of reciprocity, I offered each state a summary report of my findings in exchange for their participation. I also offered to present my findings at each state's annual MEA conference.

After reviewing the questionnaire, representatives from Florida, North Carolina, and Tennessee agreed to administer the survey study on my behalf. Upon their agreement, I sent those MEA representatives an introductory email (See Appendix B), which included a link to the online version of the study, to forward to K–12 music teachers in their state. This online questionnaire was used to solicit participating music educators' self-reports of their experiences with music teacher evaluation policy implementation, their perceptions of the teacher evaluation process, and their beliefs about evaluation policy's impact upon their daily classroom practice. I also provided state MEA representatives a follow-up email to send participants approximately 48-72 hours prior to the end of the 2-week data collection window, in order to increase response rates (See Appendix C).

After multiple attempts to contact the president and executive director from the Rhode Island MEA during late February and early March 2014, I decided to become a member of the Rhode Island MEA, as this affiliation provides members access to an online membership list. Thus, I independently solicited participation from Rhode Island MEA members, implementing the same 2-week window with a follow-up email sent approximately 48-72 hours prior to the end of data collection.

After the initial February 25<sup>th</sup> request, I sent a follow-up email to the president and executive director from Delaware on March 3<sup>rd</sup>, 2014. On March 6<sup>th</sup>, the Delaware MEA executive director informed me that, due to receiving multiple research requests similar in nature, the Delaware executive committee would discuss my request at their upcoming meeting, with the intent of reporting back to me in the next week. On March 17<sup>th</sup>, 2014, I followed up

with the Delaware MEA and received no response. Due to time constraints, I opted to drop Delaware from the study.

### **Rationale for Online Survey Methodology, Teacher Self-Report Data**

Survey methodology is frequently employed in educational research. According to Wiener and Lundy (2014), surveys in educational research have several benefits, particularly with regard to the teacher evaluation process. First, surveys capture a large amount of feedback in an efficient, low-cost manner. This feedback serves as a critical source of information as new policies are implemented. Additionally, surveys develop teachers' engagement with evaluations, and conducting surveys among educators signifies an interest and value in teachers' opinions and development.

In addition to the points outlined by Wiener and Lundy, other supporters of survey methodology note its versatility, efficiency, and generalizability. Surveys are also effective in reaching large numbers of participants in a relatively straightforward, uncomplicated manner, and they are useful in identifying patterns across various groups of participants (e.g., tenured versus non-tenured teachers). In education, surveys have been used to gather a wide array of information, including teachers' beliefs, common classroom practices, and effective teaching strategies.

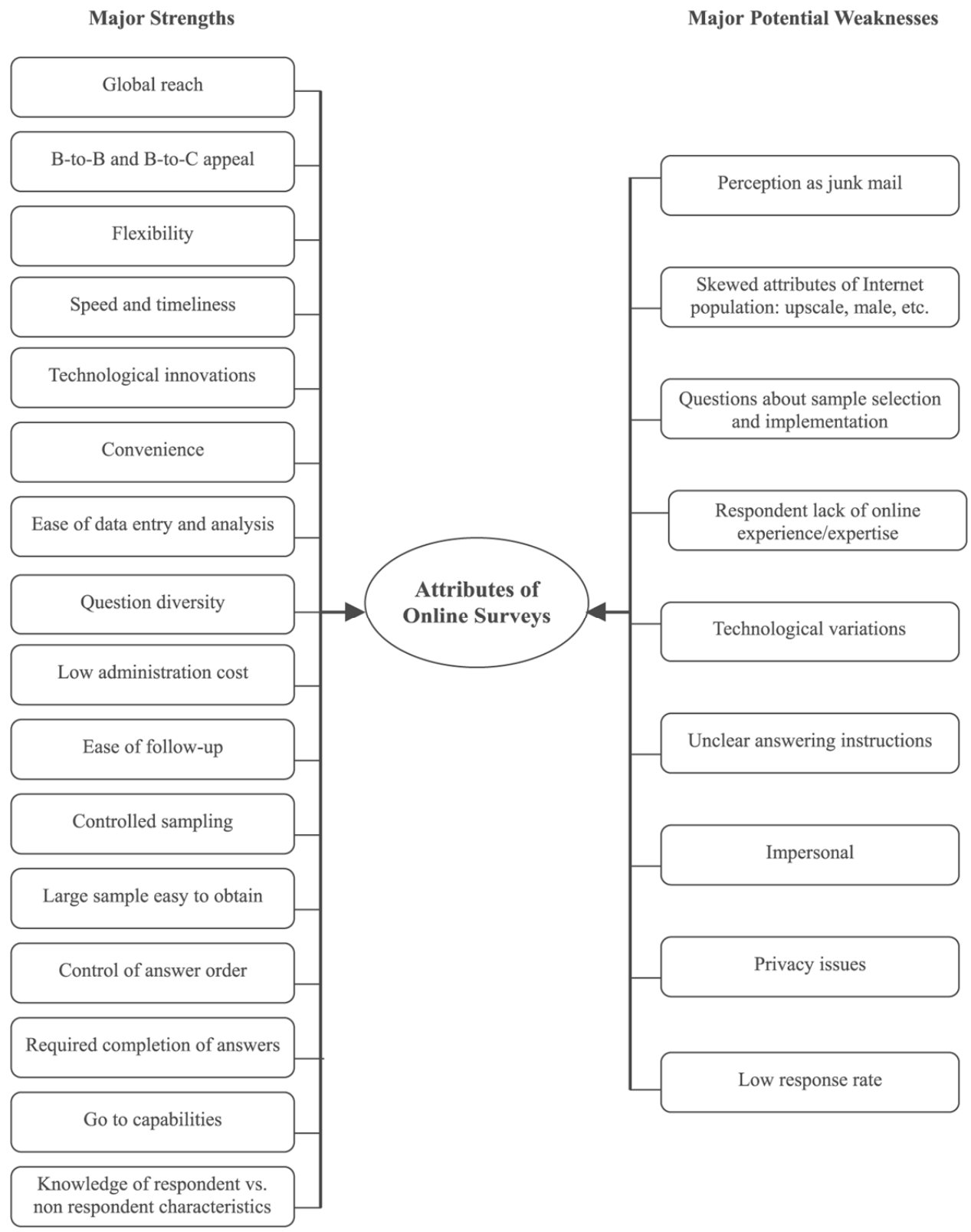
Traditionally, surveys are comprised of closed-ended items for which participants either supply or select the most appropriate response. To ensure closed-ended items effectively measure the intended constructs, items must be comprehensive and distinct from one another (Krosnick, 1999). Open-ended items employ prompts that allow participants to free-write extended responses that are not constrained to a predetermined list, thereby providing more information-rich data. Including both closed- and open-ended items in a survey instrument enables a researcher to gain a more comprehensive understanding of participant perspectives.

In this descriptive study, given my desire to ascertain the on-the-ground impact of teacher evaluation policy for music educators in multiple states, I determined that employing a survey methodology would be the most practical and efficient methodological approach for reaching a large sample size in a timely manner. To gather more comprehensive data regarding music educators' experiences with teacher evaluation, I included both closed- and open-ended items constructed to address a broad spectrum of reactions to the teacher evaluation process. Because survey methodology also supports a degree of anonymity among willing participants, responses could be considered authentic and candid.

**Online surveys.** Once a survey-based methodological approach had been selected for this study, I had to determine whether to administer a traditional paper-and-pencil questionnaire, or if an online questionnaire would be a more appropriate tool. In the case of traditional paper-and-pencil questionnaires, the target population is either offered the opportunity to complete the survey instrument in person, or the instrument is mailed to participants in the target population along with a stamped return envelope. However, in recent years, online questionnaires have become increasingly popular in survey research. These questionnaires are typically distributed to a target population through an email providing a link to the web-based survey instrument (Shih & Fan, 2008).

All methods of data collection have both strengths and weaknesses. Evans & Mathur (2005) outlined the pros and cons associated with online survey research. For example, online survey instruments have a global reach, as a large sample is easy to obtain. However, these samples are often dependent upon potential participants' access to the Internet, which can compromise sample representativeness. Many researchers appreciate the manner in which web-based survey instruments both collect and organize the data in electronic format, saving valuable time (Kaplowitz, Hadlock, & Levine, 2004), but researchers have determined paper-and-pencil

surveys yield more desirable response rates overall (e.g., Fricker & Shonlauh, 2002; Shih & Fan, 2008). Furthermore, while online survey instruments allow for a wide variety of item types (e.g., completion, forced choice, rank order, rating scales Likert-type) which may capture diverse perspectives held by members of a broad target population, the reliability of these items is heavily dependent upon item language and clarity of instructions. Figure 3.01 summarizes other advantages and disadvantages of online survey research. Ultimately, because of financial limitations and time constraints, an online survey approach emerged as the more viable methodology.



**Figure 3.01.** Strengths and weaknesses of online survey research. Reprinted from, "The Value of Online Surveys," by J.R. Evans & A. Mathur, 2005, *Internet Research*, 15(2), p. 197. Copyright 2005 Emerald Publishing Group Limited. Reprinted with permission.

**Survey platform.** There are several Internet-based online survey research platforms. Two of the most popular platforms used in educational research include SurveyMonkey and Qualtrics. While SurveyMonkey has a user-friendly interface, it is typically best implemented in basic survey research. Qualtrics, however, is intended for more complex survey designs, particularly those implemented by academic institutions. Qualtrics offers useful tools for systematic survey distribution, which simplifies the data collection process, while allowing for more item customization than is permitted in simpler platforms like SurveyMonkey. Because of its more advanced design, I chose Qualtrics as my survey platform in this study.

In creating the online version of my questionnaire, I made several design choices intended to facilitate ease of use and encourage overall completion. For example, I limited scrolling by organizing each page to have no more than three items (Ma & McCord, 2008). By including a progress bar at the foot of each page, participants were regularly updated on how far they had progressed through the survey instrument (Heerwegh, 2004; Morrel-Samuels, 2003). For the initial 12 closed-ended items soliciting participants' perspectives on the teacher evaluation process, I required answers to all questions. To connect participant responses to their respective states, I also required participants indicate the state in which they taught. All other demographic items, as well as the two open-ended questions, were optional response items.

**Teacher self-report data.** In this study, the survey instrument solicited self-report data from practicing K–12 music educators. Researchers traditionally obtain information about teaching behaviors from one of three sources: student achievement data (from which behaviors emphasized by teachers might be inferred), observer data (which are generally considered most accurate or reliable, but are difficult to obtain due to the limited availability and cost of using trained/expert observers), and self-report data (which are cost effective and generalizable, but typically considered less accurate or reliable). Efforts to validate teacher self-report data

typically involve corroborating teacher survey responses with other forms of data, including independent observer reports, student or peer reports, video analysis data, logbooks, teacher interviews, and instructional artifacts. Overall, self-report data are moderately-to-highly correlated ( $r = .40$  to  $.80$ ) with other validation criteria (Collins & Pratt, 2011; Mayer 1999).

The trustworthiness of teacher self-report data may be enhanced by basing analyses on composite scores derived from multi-item scales that provide a coherent picture of teacher beliefs or behaviors, focusing teacher responses on particular instructional contexts or subject areas and clearly delineated timeframes, and asking teachers to retrospectively report on recent events or experiences rather than projecting into the future (Koziol & Burns, 1986). While teacher self-report data may not adequately capture all qualitative aspects of classroom instruction, such data do provide fairly accurate descriptions of teacher behaviors across broad samples and contexts in ways that may inform policy development. In the context of educational reform, Rosenholtz (1987) asserts:

It makes sense to filter the effects of current reforms through the lenses of teachers involved, since only those factors that are perceived by teachers can affect their subsequent attitudes and behaviors. That is, how teachers experience policy changes will affect their commitment to them and the extent to which these interventions will have salutary effects on student learning. (p. 536)

### **Survey Responses**

In survey research, *response rate* refers to the number or percentage of responses received in relation to the questionnaires initially distributed. In the current study, the response rate of < 10% is significantly below the average online survey response rate of 33% reported by

Nulty (2008), as well as below the average online survey response rate of 36% among music teachers as reported by Miksza, Roeder, and Biggs (2012). *Non-response* refers to the number or percentage of individuals that were initially contacted who did not elect to complete the survey instrument. Survey research that has a high percentage of non-response can bias results, particularly if non-response is related to variables examined in a given survey (Hox & de Leeuw, 1994). Non-response can be a result of perceived lack of salience (Sheehan & McMillan, 1999), as Bean and Roszkowski (1995) note that participants are unlikely to respond to surveys that they find are irrelevant. Despite the timeliness of teacher evaluation policy research, it is possible that participants did not find the topic compelling, thus contributing to the low response rate in this study.

High response rates are often considered an indicator of attaining a *representative sample*. According to Ross (2005):

A sample is often described as being representative if certain percentage frequency distributions of element characteristics within the sample data are similar to corresponding distributions within the whole population. (p. 4)

The “element characteristics” described by Ross may include participants’ age, gender, or in the case of this study, area of music instruction. Connecting response rate with representativeness assumes that a high number of responses could ensure appropriate characterization of the population. However, in a review of literature on survey research, Krosnick (1999) noted that response rate does *not* beget representativeness (e.g., Visser, Krosnick, Marquette, & Curtin, 1996). In fact, Cook, Heath, & Thompson (2000) asserted, “Response representativeness is more important than response rate in survey research” (p. 821).

In this study, because state MEAs do not collect “element characteristics” regarding members’ gender, age, years in the field, or definitive information regarding their primary teaching area, determining the overall representativeness of the sample proved challenging.

Because logistical and financial constraints served to limit the potential response rate, and because sample representativeness could not be established due to state-level demographic data for NAfME members being unavailable, my primary goal was to achieve a response rate that minimized the margin of error reflected in participant responses. Given the accessible population of 6,464 music teachers and a margin of error goal of +/- 5% with a 95% confidence level, my target number of responses was 363. The margin of error for the current study was calculated to be +/- 5.65%. Therefore, despite the low response rate and uncertain representativeness of respondents, results can be interpreted with adequate confidence for the purposes of this exploratory study. The total number of participants by states is illustrated in table 3.02.

**Table 3.02**

*Total MEA Members that are K–12 Music Educators by State*

<u>State</u>	<u>N</u>	<u>n</u>
Florida	3619	81
North Carolina	1500	144
Rhode Island	145	16
Tennessee	1200	47
TOTAL	6,464	288

## Questionnaire Item Pool Development

The questionnaire developed for this study evolved from several existing measures originally designed to investigate teachers' self-reports of their experiences with and beliefs about the teacher evaluation process. Overall, the Music Teacher Evaluation Inventory (MTEI) was comprised of 23 items organized into five sections. Of these 23 items 12 were closed-ended items (soliciting participants' self-reports of the evaluation process), while two were open-ended items (asking participants to provide more detailed information about their evaluation experiences). The remaining nine items collected demographic data for each participant. The complete MTEI can be found in Appendix D.

**Preliminary study preparation: Commitment and morale.** In the earlier stages of this study, I sought to investigate the constructs of music teacher morale and commitment as impacted by teacher evaluation practices. More specifically, one of my initial research questions was, "In what ways are elements of the teacher evaluation process and/or music teachers' perceptions of the process associated with their morale and professional commitment?" I defined professional morale as, "a state of mind determined by an individual's anticipation of the extent of satisfaction of those needs perceived as significantly affecting one's total work situation" (Evans, 1997, p. 832), while commitment was conceived as the extent to which a teacher was dedicated to remaining in his or her present position. In my initial proposal, I planned to measure professional morale using the Purdue Teacher Opinionnaire created by Bentley and Rempel (1970). To measure teacher commitment, I planned to implement the 9-item Organizational Commitment Questionnaire by Mowday, Steers, and Porter (1979).

Because researchers demonstrated that so many diverse variables contribute to teacher morale and commitment, I realized it would be problematic to draw reliable correlations between the broad constructs of teacher morale and commitment with the layered concept of teacher

evaluation practices. Therefore, in revising my survey instrument, I instead explored more specific variables, such as teachers' perceived benefits of the evaluation process and their perceptions of the capabilities of their primary evaluator. In examining teachers' perceptions of these specific experiences, I could uncover, in a more direct manner, the ways in which evaluation impacted teachers on the larger scale. Furthermore, by streamlining my target variables, I was also able to establish a more efficient questionnaire, which helped foster more desirable completion times.

**Closed-ended items describing the evaluation process.** In examining related literature, I first identified existing survey instruments designed to describe practical, elemental components of the teacher evaluation process, such as who was primarily in charge of conducting teacher evaluations. To develop a comprehensive picture of teacher evaluation practices as experienced by music educators, it was critical to first understand the overall structure of each participant's evaluation system (e.g., how often are formal evaluations conducted?), as well as the types of information considered in the evaluation process (e.g., student growth data, informal vs. formal classroom observations).

Because this study encompassed evaluation processes across several states, I sought existing instruments that gathered national or multi-state data regarding the teacher evaluation process. The nuanced nature of each state's evaluation policies called for item language that could be interpreted flexibly depending on various circumstances. The majority of questionnaire items used to collect information specific to the evaluation process were adapted from the United States Department of Education National Center for Education Statistics "Survey on Teacher Performance Evaluation" (STPE) (Nolin, Rowand, & Farris, 1994). This 23-item survey instrument was administered to a nationally representative sample of approximately 1,000 public

elementary school teachers. It was originally designed to create a nationwide profile of elementary school teachers' experiences with teacher evaluation practices.

Items selected from the STPE were supported by several similar items used in the more recent Primary Sources (2013) national survey study on K–12 teachers (which included a section specific to teachers' experiences with the evaluation process) as well as items from the smaller-scale, four-state/12-district Widget Effect study on teacher evaluation (Weisberg, Sexton, Mulhern, & Keeling, 2009). The recurrence of items regarding the nature of the teacher evaluation process across multiple decades suggests a pervading interest in *how* teacher evaluation transpires, particularly in response to policy changes.

Items addressing the first research question, “What elements characterize the process used to evaluate music educators?” were primarily drawn from the STPE (Nolin et al., 1994). In the present study, participants were instructed to respond to all STPE-derived items considering their most recent teaching evaluation. For those engaged in a formal evaluation process at the time of questionnaire completion, participants responded according to the process in which they were currently engaged.

Selected items were extracted from the STPE and modified to enhance veracity. For example, in the original STPE item stem, “Which of the following were used in evaluating your teaching performance the last time you were evaluated?” was reworded to read, “For your most recent teaching evaluation, indicate the extent to which each of the following measures were considered by your evaluator(s).” The seven original STPE response items were refined to more directly reflect evaluation elements emphasized by the “Great Teachers and Leaders” recommendations. Participants indicated the degree to which measures such as formal classroom observations, professional development activities, and student achievement data informed their

overall evaluation. In addition, to reflect a music-specific context, participants also indicated the degree to which student performances/contest ratings factored into their overall evaluation.

While the STPE relied on a yes-or-no format, I adopted a 4-point rating scale format, where 1 = *not at all considered/does not apply*, 2 = *slightly considered*, 3 = *considered to a moderate extent*, and 4 = *considered to a great extent*. The rating scale format was utilized to better understand not only what measures were considered, but also how much weight each measure carried in the overall evaluation process. Participants were also given the option to indicate, “I am not sure if this was considered,” to account for a potential lack of clarity surrounding certain components of the evaluation process and minimized the influence of response error.

Participants could also identify unique measures of teaching considered in their evaluation process that were not listed as options in the questionnaire. For those who indicated that student achievement data were considered in their overall evaluation, a subsequent selection-type item appeared in the questionnaire, asking participants to indicate the types of student achievement data considered in their teaching evaluation. Options included non-music standardized test score data, state or district developed music-specific assessments, and portfolio systems examining student work samples. Participants then indicated the extent to which they believed each measure *should be* considered in their evaluations and were offered the opportunity to list alternative measures of teaching they felt should be a part of their overall evaluation.

The next section of the questionnaire addressed evaluative objectives. On the original STPE, participants indicated whether their most recent evaluation was conducted as part of a regularly scheduled review, and whether the results were used to determine tenure, promotion or merit pay. For the current study, participants considered the extent to which various objectives

were weighted in the evaluation process. For example, participants indicated the degree to which they believed teacher evaluations were used to “discharge incompetent teachers” or “guide improvement of teachers’ skills.” Participants responded using a 4-point rating scale, where 1 = *not considered an objective at all*, 2 = *to a small extent, this is an objective*, 3 = *to a moderate extent, this is an objective*, and 4 = *to a great extent, this is an objective*. Responses signifying, “I’m not sure if this is considered an objective,” were treated as missing data. As with the previous questionnaire section, participants were also offered the opportunity to identify unique objectives, and indicate the extent to which they believed each objective *should be* considered as part of their evaluations.

In addition to soliciting information regarding the measures by which participants were evaluated and the objectives of the evaluation process, several items focused upon evaluators themselves. The STPE item, “Who was involved in evaluating your teaching performance the last time you were evaluated?” initially listed options such as “your principal,” “master teacher,” and “school board.” On the original STPE survey, participants responded to each option using a yes-or-no format. To reflect the importance of evaluator training and expertise in determining music teacher effectiveness, modified options for identifying major evaluators included “building administrator with NO music teaching expertise,” and “building administrator WITH music teaching expertise.” Other options included (a) a district music or fine arts administrator, (b), other music teacher(s), and (c) other teacher(s) (non-music). Participants were also able to cite other individuals who may have been involved in their evaluative process. Similar to the STPE, participants were then asked indicate which individual had the *most authority* in determining their overall evaluation.

To determine how teachers received (or expected to receive) feedback upon the completion of their formal evaluation, the original STPE asked participants whether they

received a written report or a verbal explanation of their evaluation, using a yes-or-no response format. For this study, additional evaluation result dissemination options were included, such as “a conference with your primary evaluator to review your evaluation.” Participants indicated all forms of feedback they received or expected to receive upon completing the evaluation process.

**Closed-ended items on teacher perceptions.** Researchers investigating teachers’ perceptions of evaluation systems have considered a range of beliefs and constructs, including impressions of fairness, clarity, utility, control, and overall support for the evaluation process, as well as evaluator credibility and competence (Kelly, Ang, Chong, & Hu, 2007; O’Pry & Schumacher, 2012; Palazuelos, 2007). Through examining these studies, I narrowed my focus to the following constructs:

1. Teacher perceptions of the overall evaluation process (i.e., fairness, clarity, utility)
2. Teacher perceptions of the primary evaluator (i.e., objectivity, experience, qualifications)
3. Teacher perceptions of evaluation’s benefit to their professional practice (i.e., improved effectiveness, increased reflective practice, improved student learning)
4. Teacher perceptions of evaluation’s impact on daily classroom practice (i.e., influence on material choice, impact upon instructional methodologies, effect on instructional time)

The majority of items gauging music teachers’ perceptions of various aspects of the teacher evaluation process were derived from Kelly et al.’s (2007) instrument designed to measure teachers’ impressions of the appraisal system in Singapore. Additional items were drawn from Schumacher’s (2010) and O’Pry and Schumacher’s (2012) studies on teacher

perceptions of evaluation systems. Items from these measures were modified to enhance item clarity and validity. For example, the Kelly et al. item, “The appraisal system is transparent,” was modified to read, “The teacher evaluation process is clear to me.” The Schumacher (2010) item, “The feedback that I receive from my evaluator is an accurate reflection of my teaching performance,” was modified to read, “Our evaluation process provides an accurate depiction of my teaching performance.” Participants responded to each of the overall perception items on a 5-point, Likert-type scale, where 1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, and 5 = *strongly agree*.

In addition to the descriptive information gathered about each participant’s primary evaluator in the STPE-adapted section of the questionnaire, several items addressed teachers’ perceptions of their evaluator’s qualifications and expertise. For the purposes of this study, all language reflecting the term “appraiser” as used in the Kelly et al. measure was modified to read “evaluator.” For example, the Kelly et al. item, “I have a good interpersonal relationship with my appraiser,” was amended to read, “I have a good interpersonal relationship with my evaluator.” As with prior items, a 5-point Likert-type scale was used to measure responses. An optional, open-ended item was also included, asking participants to describe any other factors about their primary evaluator they felt relevant.

Participants were asked to reflect upon the potential benefits of the evaluation process. More specifically, participants indicated the extent to which they agreed or disagreed with the evaluation process impacting various aspects of their professional development and student achievement. For example, the O’Pry and Schumacher (2012) item, “My overall evaluation helped me improve my teaching practice,” was modified to read, “The evaluation process has helped me become a more effective teacher.”

Participants were also asked to report how the teacher evaluation process impacted various elements of their daily teaching practice, such as the ability to assess students in a manner relevant to the content area, as well as the ability to utilize instructional time in a manner that best benefitted students. Participants responded to these researcher-developed items on a 5-point bi-polar scale, ranging from *extremely negative impact* to *extremely positive impact*.

**Open-ended items.** In addition to the 12 closed-ended items focusing upon music educators' perceptions of teacher evaluation, two open-ended items solicited participants' detailed reflection upon more specific aspects of teacher evaluation's impact upon their practice and beliefs. In survey research, open-ended items refer to stems that do not provide predetermined responses; rather, participants are responsible for responding to a given prompt or question in their own words (Ballou, 2011). Open-ended items are included in survey research to "garner longer and richer responses" that "solicit a reflection, an explanation of a behavior, or a reaction to a situation" (Jones, Torres, & Arminio, 2006, p. 147). Including open-ended items encourages participants' spontaneous and authentic response, as closed items might limit or bias participants by suggesting specific answers (Reja, Manfreda, Hlebec, & Vehovar, 2003). Schuman and Presser (1996) suggest that when participants are asked to come up with answers on their own, more valid data are gathered.

While open-ended items can enhance understanding about a given construct or phenomena, there are challenges to including open-ended items in survey research. First, open-ended items result in non-response more often than closed-ended items. Participants' limited understanding of the question or lack of comfort with self-expression can both contribute to non-response. Open-ended items can also be more difficult to interpret, analyze, and validate (Geer, 1991; Roberts et al., 2014; Schuman & Presser, 1996). Despite these challenges, including open-ended responses in survey research can greatly enhance researchers' understanding of participant

perspectives. In fact, in music education survey research, open-ended items are commonly used to support closed-ended data gathered from practicing K–12 music teachers (Legette, 2003, Miksza et al., 2012; Sindberg & Lipscomb, 2005).

Ballou (2011) describes the rationale for various types of open-ended items. For example, open-ended items can be used to expand a list by including a follow-up to encourage participants to describe a response that may not have been provided in an initial closed-ended item. This approach can include a prompt such as “Other (please specify).” Another type of open-ended item might ask participants to explain a previous answer by providing additional details (e.g., “Justify your response to the previous item”). A third type of open-ended item might explore a new subject all together in a long-response format, which can offer insights into participants’ theoretical grounding in previous responses or highlight salient topics not otherwise covered in closed-ended items. Including a variety of open-ended items throughout a survey instrument can build rapport with the participant by signifying their opinion is valued.

***Open-ended item development.*** In the earliest stages of constructing the survey instrument for this study, long-response open-ended items were not included in the questionnaire. The only open-ended items included served the purpose of list expansion, allowing for participants to contribute other types of information not initially provided in closed-ended options. For example, in the initial version of the questionnaire, participants were offered an “Other (please specify):” prompt following a question regarding the types of individuals involved in their evaluation process.

Throughout the various stages of survey instrument refinement, it became apparent that a long-response open-ended item might offer support for participants’ responses to the closed-ended item regarding evaluation’s impact on overall teaching effectiveness. In order to contextualize this topic for the reader, the prompt was initially written as follows:

Many teacher evaluation processes are designed to develop more effective teachers. One manner in which many gauge effective teaching is through increased student learning. Considering your own experiences, do you feel the teacher evaluation process at your school helps develop your effectiveness as a music teacher? Why or why not?

In the first pilot testing process, experienced music teachers provided a wide variety of responses that indicated the original prompt lacked clarity and direction. To facilitate more focused responses, the prompt was revised for clarity and streamlined for brevity. Because the primary purpose of teacher evaluation is to ensure highly effective educators are actively contributing to student learning, the final prompt offered participants an opportunity to elaborate upon their closed-ended responses to the “teacher perceptions of evaluation’s benefits” scale. The revised prompt read, “Describe how the teacher evaluation process at your school impacts music student learning and/or your effectiveness as a music teacher.”

An additional long-response open-ended item was included to offer participants the opportunity to reflect upon teacher evaluation’s impact upon their feelings of professionalism. The initial version of this prompt was rooted in Crowe’s (2008) concept of requirements for attaining professional legitimacy, including “shared norms, training, working practices, and regulatory mechanisms” (p. 990). These requirements informed the terminology of “quality control” used in the early version of this open-ended prompt:

The concept of “quality control” is embedded in many professions. People such as doctors, lawyers, and engineers are held to certain performance standards as part of their profession. Consider music teaching as a profession, and reflect on whether you feel the teacher evaluation process is an effective means of “quality control.” Does the teacher evaluation process make you feel like more of a professional? Why or why not?

Again, during the first round of pilot testing, experienced music educators’ responses reflected confusion regarding the question’s intended direction. One participant responded, “Yes, I think there should be an evaluation process, but it should measure what it’s supposed to measure.” Another participant remarked, “I think that most professions have some sort of evaluation process, even if they work at McDonald’s.” To enhance item clarity for the second open-ended prompt, the revised version read:

How has the teacher evaluation process impacted your perception of music teaching as a *profession*? Does this process elevate or diminish the professional status of music teachers?

The revised versions of each prompt were implemented in the second round of pilot testing with two additional experienced K–12 music educators. Participants’ responses were more reflective of each prompt’s intent. Therefore, both of the simplified versions of the long-response open-ended items described above were used in the final survey instrument.

**Demographic data.** Finally, demographic information, including participant gender, area of specialty (e.g., general music, band, orchestra, vocal music), and years of teaching experience, was collected. Because NAfME membership is not restricted to practicing K–12 music teachers (i.e., private instructors or teachers at institutions of higher education may also become members), one item addressed the nature of each participant’s teaching position.

## **Procedures**

**Pilot testing and pre-data collection.** Upon receiving approval from the University of Colorado Institutional Review Board (See Appendix E), an initial version of the online MTEI (26 items) was constructed and pilot tested with six practicing K–12 music educators in a western suburb of Denver, Colorado in January 2014. These pilot test candidates were identified due to their familiarity with the implementation of a new teacher evaluation system. In 2010, Colorado passed Senate Bill 10-191, also known as the Colorado Model Evaluation system, in its effort to meet the requirements of Race to the Top grant application category, “Great Teachers and Leaders.” As per the legislation, a state model for teacher evaluation was developed through the Colorado Department of Education. Local education agencies were afforded the option of implementing the provided system, or designing their own teacher evaluation system from the state model. This evaluation system was in its first year of full implementation during the 2013–2014 school year.

The pilot questionnaire was designed and administered through Qualtrics online survey software. In January of 2014, pilot participants were asked to complete the online questionnaire and provide feedback regarding item clarity, as well as provide any other recommendations for improving the quality and efficiency of the instrument. The Qualtrics system automatically logged each participant’s questionnaire completion time. Among the six pilot participants,

average questionnaire completion time was 15 minutes and 56 seconds. The pilot version of the questionnaire can be found in Appendix F.

Based upon written feedback from the pilot participants, some items considered unnecessarily redundant were eliminated, and additional open-ended response options were created for specific items so as to allow main study participants the opportunity to articulate additional evaluative approaches/outcomes beyond overt and obvious choices. Furthermore, upon reviewing the varied responses pilot study participants provided to the open-ended, short-answer items, it was apparent the prompts for these items were unclear. As a result, narrative prompts associated with the open-ended items were refined to enhance item clarity and better focus responses.

Following the initial pilot test, a revised version of the questionnaire was then administered to two additional experienced music educators in Colorado. Subsequent feedback resulted in the decision to include a “does not apply” option for general music teachers, when considering whether or not contest ratings informed the student growth component of their evaluation. Because average completion time continued to exceed 15 minutes, my advisor and I revised item language and structure to reduce the average completion time. The final 23-item version of the questionnaire took approximately 12 minutes to complete. After the instrument was finalized, state invitation emails and participant email scripts were refined and personalized to address each particular state. Representatives from each state MEA were contacted via email, and these representatives were provided with the survey instrument and participant email scripts to determine their interest in inviting state MEA members to participate in the study.

**Data collection.** Music teachers in the four participating states who were NAFME members were contacted via email by their state MEA office. The email script in the initial invitation email (See Appendix B) described the nature of the study and provided participants a

link to the questionnaire itself. The online questionnaire began with an informed consent form. All MTEI items collected information regarding music teachers' evaluation experiences, as well as their perceptions about the clarity, fairness, and value of the evaluation process. All data were collected throughout a 12- to 14-day window during March of 2014. Participants were sent a final reminder email approximately 48 hours prior to the end of data collection window.

### **Data Analysis**

Data analysis began with descriptive analyses of music teachers' evaluation processes. Frequencies and percentages were calculated for all categorical variables related to the teacher evaluation process, and means standard deviations were determined for all continuous variables. Due to the varied response rates across the four target states, descriptive statistics were reported in aggregate form. Reliability analyses were conducted to determine the reliability (internal consistency as determined by coefficient alpha) for each multi-item subscale included within the questionnaire, and paired samples *t*-tests were used to examine participants' lived and idealized notions of certain evaluation characteristics. Group comparisons were used to determine mean differences in teacher perceptions across various individual difference variables, including gender, teaching area, and teaching level. For open-ended items, descriptive and in vivo coding was conducted to determine if patterns existed with regard to participants' beliefs about the impact of the evaluation process (Saldana, 2012). For more detailed information on open-ended item coding and analysis, see page 127 of Chapter 4.

## Chapter IV

### Results

#### Introduction

Nearly 300 practicing K–12 music educators, representing four states (Florida, North Carolina, Rhode Island, Tennessee) that had received Race to the Top funding and implemented a new teacher evaluation system within the past two years, participated in an exploratory online survey study. The survey instrument was a 23-item questionnaire that included a combination of checklist items, rating scale items, and open-ended response items. These items, adapted in part from existing measures developed by the United States Department of Education’s National Center for Education Statistics and other researchers (Kelly et al., 2007; O’Pry & Schumacher 2012), were designed to collect self-report data regarding current teacher evaluation policies implemented at the music classroom level, how music teachers perceive these policies in terms of clarity, fairness, and utility, and the degree to which music teachers may be adapting their professional decision making and practice in response to these policies. Participants spent an average of 21 minutes completing the questionnaire.

Participant responses to checklist and rating scale items were analyzed using the Statistical Package for Social Sciences (SPSS, Version 22.0.0.0, for Mac OSX, 2013). These results are summarized in Section A of this chapter. Basic descriptive analyses were conducted to summarize participant demographic data. All subsequent results are organized according to questionnaire subsections. Each subsection summary first includes descriptive information, followed by inferential analyses. Paired samples *t*-tests were conducted to determine the difference between participants’ actual evaluation experiences in terms of evaluative objectives and criteria that *are* considered to a certain extent, and what they believed *should* be considered part of their evaluation experiences. Group differences in music educators’ perceptions of

teacher evaluation as related to factors such as instructional area (e.g., general music, band, orchestra, choir) and instructional level (e.g., K-5, 6-8, 9-12) were examined using multivariate analyses of variance (MANOVA).

The online questionnaire also included two major open-ended items. Responses to these open-ended items were coded inductively and organized according to theme. An expert external audit was employed to ensure code veracity and check for overall coding consistency. Results for open-ended items are summarized in Section B of this chapter.

While a total of 438 participants began the online questionnaire, only 68% completed it entirely. Participant attrition is reflected in the results section, as descriptive statistics are reported according to the number of participants that completed each questionnaire subsection. Group comparisons are based on the total number of participants that completed the full questionnaire.

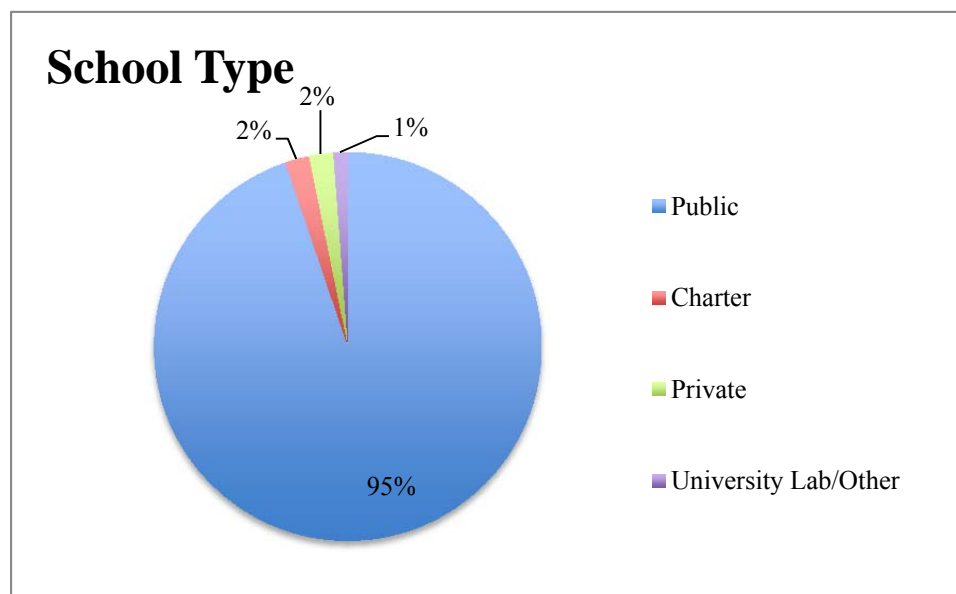
## **Section A: Closed-Ended Items**

### **Participant Demographics**

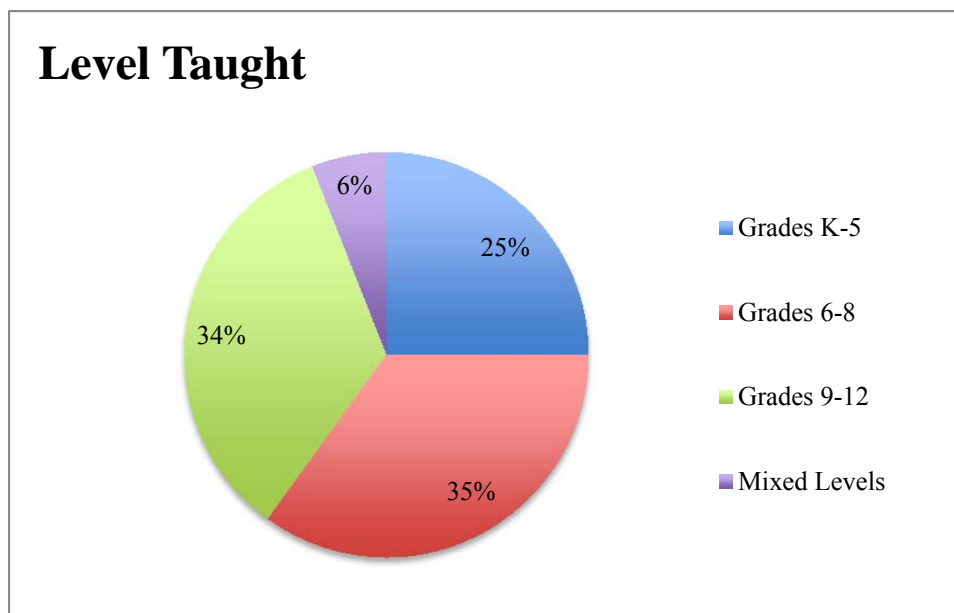
A total of 6,464 practicing K–12 music educators were invited to participate in the study. Of those invited, a total of 299 participants fully completed the online questionnaire. Participants were asked to indicate the level at which they spent the majority of their time teaching. A total of three participants reported teaching primarily at the collegiate level. Because this study focused upon practicing K–12 music teachers, these three participants were removed from consideration for all subsequent analyses. A separate item asked participants to indicate the most recent school year in which they participated in a formal teaching evaluation. A total of eight participants indicated the most recent time they were formally evaluated was prior to the 2012–2013 school year. Because this study examines perceptions of teacher evaluation policy implemented during or after the 2012–2013 school year, these participants were also excluded from subsequent

analyses. Therefore, the total number of eligible participants that completed the entire questionnaire was 288, resulting in a margin of error of  $\pm 5.65\%$  with a 95% confidence level. Descriptive statistics are reported according to the number of participants that completed each questionnaire subsection. For example, though 288 participants completed the entire questionnaire, a higher number of participants completed earlier portions of the survey instrument. Therefore, the total  $n$  will vary throughout the results section.

Participants included 155 females (54.2%) and 126 males (44.1%). A total of seven participants chose not to respond to the gender item. The majority of participants indicated they had tenure status (71.9%) and for the vast majority of participants ( $n = 281$ ), music represented 50% or more of their teaching load. Participants reported an average of 15.97 years of teaching experience ( $SD = 9.21$ , range: 1–41). Almost all participants (95%) taught in public schools (see Figure 4.01), and most participants (75%) taught at the secondary level or a combination of elementary and secondary levels (see Figure 4.02).

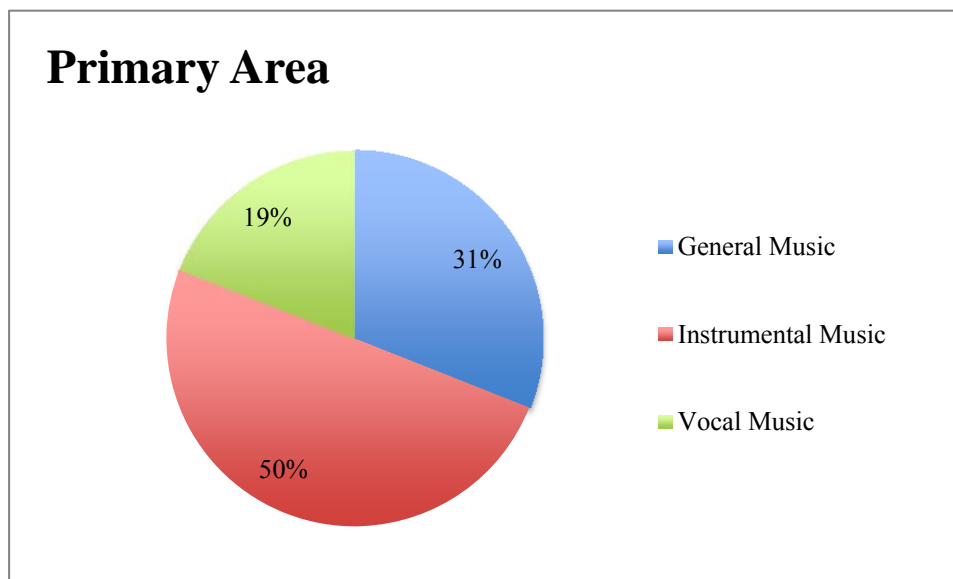


**Figure 4.01.** Type of school in which study participants taught music ( $n = 288$ ).



**Figure 4.02.** Participants' primary teaching level ( $n = 288$ ).

To help facilitate comparisons across groups, participants were also asked to indicate the music education area in which they spent the majority of their time teaching (Figure 4.03). Responses were aggregated into three main categories: (a) general music (including theory, guitar, and piano classes), (b) instrumental music (including band, orchestra, and jazz), and (c) vocal music (including musical theater). The majority of participants primarily taught instrumental music ( $n = 143$ ), while 88 participants primarily taught general music, and 54 participants primarily taught vocal music.



*Figure 4.03.* Participants' primary area of teaching ( $n = 288$ ).

### Characteristics of the Evaluation Process

One of the primary purposes of this study was to identify and describe various characteristics of the evaluation process, as seen from the perspective of practicing K–12 music educators. Results in this section therefore reflect music teacher self-reported information regarding the overall purpose of teacher evaluation, measures of teaching performance considered in the evaluation process, and evaluator characteristics. All descriptive analyses represent total responses across all participating states.

**Evaluative purpose.** To determine how various evaluation systems functioned within school contexts, participants weighed a variety of professional outcomes in relation to their personal evaluation experiences. Using a 4-point rating scale ranging from one (not at all considered/does not apply) to four (considered to a great extent), a total of 296 participants completed items regarding the purpose(s) of their evaluation system. Participants were asked to

indicate the extent to which certain outcomes *were* considered objectives in their most recent teacher evaluation. Participants were also asked to indicate the extent to which certain outcomes *should be* considered objectives in there evaluations.

Participants reported that “guiding improvement of teachers’ skills” was the most heavily emphasized objective in the evaluation process ( $M = 3.06, SD = 0.79$ ), while “merit pay” was least considered ( $M = 1.99, SD = 1.17$ ). “Guiding improvement of teachers’ skills” was also identified as the objective that *should be* most strongly considered ( $M = 3.74, SD = 0.54$ ), while participants felt “merit pay” *should be* considered only slightly as an evaluation purpose ( $M = 1.75, SD = 0.94$ ). Overall, participants responded more favorably to objectives associated with professional development (e.g., to reinforce excellent teaching), and were less supportive of objectives emphasizing professional status outcomes (e.g., to make tenure decisions), with regard to both *what is* and *what should be*. Means and standard deviations for evaluation objectives are summarized in Table 4.01.

**Table 4.01***Means and Standard Deviations for Evaluation Purpose (n = 296)*

<b>Purpose</b>	<b><i>Is considered</i></b>		<b><i>Should be considered</i></b>	
	<b><i>M</i></b>	<b><i>SD</i></b>	<b><i>M</i></b>	<b><i>SD</i></b>
To guide improvement of teachers' skills	3.06	0.79	3.75	0.54
To help teachers better focus upon student outcomes and growth	2.95	0.87	3.48	0.68
To recognize and reinforce excellent teaching	2.88	0.92	3.67	0.60
To discharge incompetent teachers	2.42	0.93	2.84	0.88
To make tenure and promotion decisions	2.10	1.07	2.53	0.97
To determine teachers' merit pay	1.99	1.17	1.81	0.97

*Note.* 1 = This is not/should not be considered, 2 = This is/should be considered slightly, 3 = This is/should be considered to a moderate extent, 4 = This is/should be considered to a great extent.

A series of paired samples *t*-tests were used to compare teacher reports of *what is* and *what should be* with regard to the degree of emphasis placed on various evaluative objectives. A Bonferroni adjusted significance level of .008 (.05/6) was implemented to reduce the chance of Type I error. There was no significant difference between participants' average *what is* and *what should be* responses for the objectives "to determine teachers' merit pay". Results for all other objectives demonstrated statistically significant differences ( $p < .001$ ), reflecting music teachers' desire for several evaluative functions to be considered to a greater extent. Effect size, which refers to the magnitude of this significance, was calculated using Cohen's *d*. Effect sizes for "to

guide improvement of teachers' skills" and "to recognize and reinforce excellent teaching" were large ( $d > .80$ ) (Cohen, 1988). Table 4.02 summarizes results from the paired samples  $t$ -tests.

**Table 4.02**

*Paired Samples  $t$ -Test for Evaluation Purpose ( $n = 296$ )*

Variable	Paired Differences		$t$	$df$	Sig.	$d$
	$M$	$SD$				
To guide improvement of teachers' skills	-0.69	0.85	-13.94	295	< .001	1.00
To help teachers better focus upon student outcomes and growth	-0.53	0.93	-9.68	294	< .001	.68
To recognize and reinforce excellent teaching	-0.78	0.99	-13.39	289	< .001	1.02
To discharge incompetent teachers	-0.48	1.14	-6.64	246	< .001	.46
To make tenure and promotion decisions	0.25	1.43	3.06	252	< .001	.39
To determine teachers' merit pay	0.19	1.39	2.16	256	.032	.23

*Note.* Adjusted significance level of .008.

If participants were unsure about a given purpose being an objective of their evaluation process, they had the option of indicating, "I'm not sure if this was considered an objective." Such responses were considered missing data for the purposes of comparative analyses. However, a notable proportion of participants (13%,  $n = 42$ ) were unsure if the evaluation process was used in informing the discharge of incompetent teachers. Nearly 40% of participants

( $n = 133$ ) were unclear as to whether or not their evaluations informed merit pay decisions, and another 30% of participants ( $n = 99$ ) were not sure if their evaluations contributed to tenure and promotion decisions.

**Measures of teacher performance.** Participants were asked to indicate the extent to which various measures of teacher performance were considered in their most recent evaluation, as well as the extent to which these same measures *should be* considered. Item formats and response options were identical to those used to collect music teacher reports of evaluative objectives.

Overall, participants indicated that formal classroom observations were considered most heavily in the evaluation process ( $M = 3.51$ ,  $SD = 0.79$ ), while student input/feedback was considered the least ( $M = 1.69$ ,  $SD = 0.93$ ). Similarly, participants indicated that formal classroom observations *should be* considered heavily in the evaluation process ( $M = 3.54$ ,  $SD = 0.58$ ), while student input/feedback *should be* considered least ( $M = 2.31$ ,  $SD = 0.93$ ). When asked if any other measures of teaching were considered in participants' most recent evaluation, responses included "extracurricular performance expectations" ( $n = 2$ ) "documentation of communication with parents" ( $n = 1$ ), and "classroom management" ( $n = 1$ ). Means and standard deviations for measures of teaching considered in the evaluation process are summarized in Table 4.03.

**Table 4.03***Means and Standard Deviations for Measures of Teacher Performance (n = 320)*

<b>Measures of Teacher Performance</b>	<b>Measures are considered</b>		<b>Measures should be considered</b>	
	<b>M</b>	<b>SD</b>	<b>M</b>	<b>SD</b>
Formal classroom observations	3.51	0.79	3.54	0.58
Informal classroom observations	3.04	0.90	3.34	0.62
Professional development activities	2.43	0.97	2.87	0.80
Student achievement data (growth)	2.34	1.17	2.57	0.94
Student performances/contest ratings	1.83	1.05	2.67	1.00
Portfolio of your work	1.76	1.00	2.81	0.98
Student input/feedback	1.69	0.93	2.31	0.93

*Note.* 1 = This is not/shout not be considered, 2 = This is/should be considered slightly, 3 = This is/should be considered to a moderate extent, 4 = This is/should be considered to a great extent.

A series of paired-samples *t*-test (Bonferroni adjusted significance level of .007) were used to compare music teachers' "what is" and "what should be" responses for evaluation sources. Responses differed significantly for all measures except formal observations,  $t(319) = -1.50, p = 1.35$ . These results suggest that participants' believe most measures *should be* considered to a greater extent than they currently are. Results from the paired samples *t*-tests are summarized in Table 4.04.

**Table 4.04***Paired Samples t-Test for Measures of Teacher Performance (n = 320)*

<b>Variable</b>	<b>Paired Differences</b>			<b>df</b>	<b>Sig.</b>	<b>d</b>
	<b>M</b>	<b>SD</b>	<b>t</b>			
Formal classroom observations	-.08	0.90	-1.50	319	.135	1.02
Informal classroom observations	-.40	1.10	-6.50	319	< .001	.78
Professional development activities	-.58	1.32	-7.83	319	< .001	.56
Student achievement data (growth)	-.40	1.49	-4.86	319	< .001	.45
Student performances/contest ratings	-1.03	1.33	-13.87	319	< .001	.24
Portfolio of your work	-1.23	1.34	-16.44	319	< .001	.23
Student input/feedback	-.91	1.17	-13.97	319	< .001	.19

*Note.* Adjusted alpha level of .007.

Nearly one-third of participants ( $n = 101$ ) indicated that student achievement data were “not at all considered” during their most recent teaching evaluation. This finding is particularly notable considering student growth is a principal component of teacher evaluation policy set forth by the Race to the Top grant program. However, participants from North Carolina ( $n = 144$ ) were not yet subject to student growth playing a role in their overall evaluations. Still, only 66 of the 101 participants that indicated student achievement data were “not at all considered” were from North Carolina.

Participants reporting that student achievement data were considered as part of their most recent evaluation ( $n = 202$ ) also could identify the types of growth data used from a list of typical sources of student achievement data. A total of 94 participants revealed that the student

achievement data component of their evaluation involved non-music standardized test score data for students they *did not* teach (e.g., building-level growth data), while an additional 92 participants indicated the use of non-music standardized test scores of those students they *did* teach. A total of 31 participants reported being evaluated on the basis of state- or district-developed standardized paper-and-pencil tests specific to music, while 30 participants referenced state- or district-developed music performance tasks. An additional 33 participants indicated that a portfolio-based system, designed to examine samples of music students' work, was part of the student achievement data evaluative component. In interpreting these varied responses, it is important to note that the participating states were each pursuing

An open-ended item allowed participants to list additional sources of student growth that were considered in their evaluations. Other sources mentioned included teacher-developed music assessments (i.e., created at the classroom level rather than the district or state level,  $n = 11$ ) and teacher-developed Student Learning Objectives (SLOs,  $n = 5$ ). Frequencies for types of student achievement data considered in music teacher evaluations are summarized in Table 4.05.

**Table 4.05**

*Frequencies for Types of Student Achievement Data Considered in Music Teacher Evaluations  
(n = 334)*

<b>Role</b>	<b>Frequency</b>	<b>% total n</b>
Non-music standardized test score data for students I do NOT teach	98	29%
Non-music standardized test score data for students I DO teach	92	28%
Portfolio based system designed to measure music student growth	37	11%
State or district paper-and-pencil test designed to measure music student growth	33	10%
State or district performance task designed to measure music student growth	32	10%
Teacher-developed classroom level music-specific assessments	11	3%
Teacher-developed Student Learning Objectives	5	1%
Performances as indicators of growth	4	1%
AP Music Theory Exam	1	< 1%

*Note.* Participants were able to indicate more than one type of student achievement data considered in their evaluation process.

**Evaluative personnel.** A total of 334 participants completed items about the professional role(s) of their evaluator(s) (e.g., building administrator, district or fine arts supervisor, other music teachers). Music teachers were asked to indicate all of the individuals that were involved in and/or had input on their most recent teaching evaluation. Only 13% of participants had more than one individual involved in their teaching evaluation ( $n = 43$ ). Response frequencies for evaluator role are summarized in Table 4.06. In response to an open-ended item soliciting the professional roles of additional individuals involved in the evaluation process, several

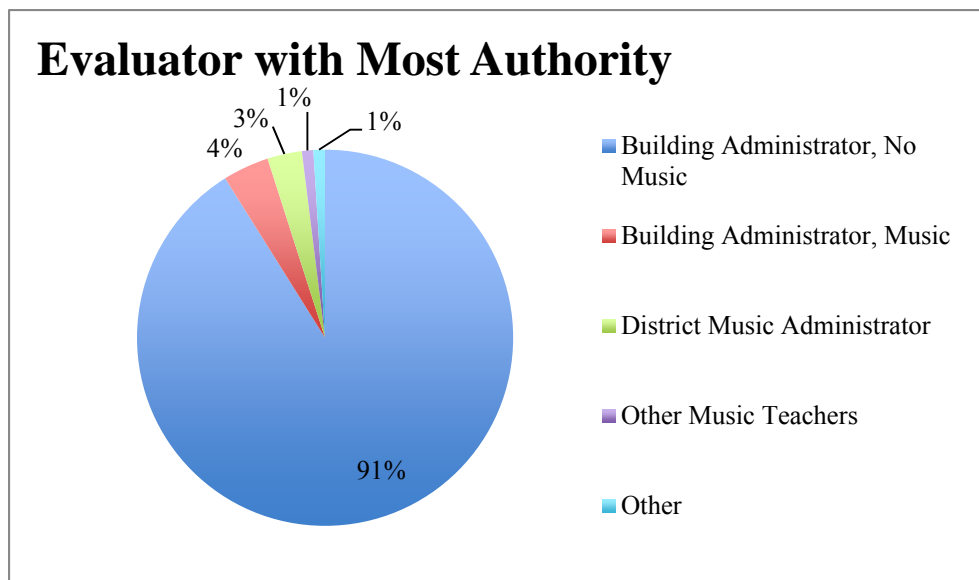
participants cited non-music district-level curriculum coordinators and administrators with limited experience in music. An additional item asked participants to report which individual had the most authority in determining their overall evaluation. Approximately 92% of participants indicated their primary evaluator had no music expertise or background ( $n = 291$ ). Figure 4.04 displays the distribution of responses regarding what types of evaluators had the most authority in participants' evaluations.

**Table 4.06**

*Frequencies for People Involved in Music Teacher Evaluation (n = 334)*

<b>Role</b>	<b>Frequency</b>	<b>% total <i>n</i></b>
Building administrator with no music teaching experience	303	91%
Building administrator with music teaching experience	19	6%
District music or fine arts administrator, supervisor, or coordinator	8	2%
Other music teacher(s)	10	3%
Other teacher(s) (non-music)	24	7%

*Note.* Participants were able to indicate more than one type of person involved in their evaluation process.



**Figure 4.04.** Evaluator with most authority in music teacher evaluation ( $n = 332$ )

In addition to describing the professional role(s) of their evaluator(s), participants shared their opinions regarding the overall approach and competence of their primary evaluator in administering teacher evaluations by responding to six Likert-type items, with response options ranging from one (*strongly disagree*) to five (*strongly agree*). Responses suggested a generally neutral position on the primary evaluator's overall approach and competence in administering teacher evaluations. Participants responded most favorably to having a strong interpersonal relationship with their evaluator ( $M = 4.03$ ,  $SD = 9.90$ ), and participants were least confident in their primary evaluator's qualifications to assess their teaching ( $M = 3.29$ ,  $SD = 1.16$ ). Table 4.07 summarizes results from the evaluator competence scale.

**Table 4.07**

*Means and Standard Deviations for Participants' Beliefs about Primary Evaluator's Competence (n = 313)*

<b>Competency</b>	<b>M</b>	<b>SD</b>
I have a good interpersonal relationship with my evaluator.	4.03	0.99
My evaluator is fair and objective.	3.88	1.08
My evaluator has considerable experience in teaching.	3.73	1.06
My evaluator does a thorough job.	3.65	1.08
I trust my evaluator to accurately assess my teaching performance.	3.38	1.16
My evaluator is qualified to assess my teaching.	3.29	1.16

*Note.* 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree.

A composite rating for evaluator competence was calculated by summing individual responses to each of the six competencies presented in Table 4.07. Reliability estimates for the composite ratings (responses summed across all six items) of evaluator competence were strong ( $\alpha = .91$ ). The mean overall score was 21.96 ( $SD = 5.45$ ), with a possible range of 6 (e.g., evaluator displaying little competence) to 36 (evaluator displaying exemplary competence).

This result, which suggests participants are only marginally secure in their primary evaluator's overall approach and qualifications, was supported in part by participants' responses to the brief open-ended prompt, "Please describe any other factors about your primary evaluator that you believe are relevant." A total of 78 participants contributed additional thoughts about their primary evaluator. Responses were coded and organized by theme. The majority of comments focused upon primary evaluators who lacked music expertise or training. For

example, one participant noted, “My evaluator has no formal training in music education and does not really understand the essential standards,” while another participant wrote:

None of my administrators are music people, and sometimes they don't get that what I do is part of the evaluation, [it] just looks different in a band room. [For example] – cooperative learning. Band is the ultimate cooperative learning [experience], but they don't see it that way.

Though the concern about primary evaluators’ lack of musical expertise recurred through participants’ responses, many participants also acknowledged that their primary evaluator still strived to be fair and equitable in the evaluation process. One participant contributed, “My evaluator does not know much about music education, but he does ask great questions that help him understand my content,” and another stated, “Though [my primary evaluator] doesn’t know music, she listens to my needs and supports me in every way.” Still other participants noted that their primary evaluator was chiefly concerned with the quality of school music performances. As one participant observed, “As long as the concert sounds good, [my evaluators] leave me alone.”

Other themes that emerged surrounded the politics of evaluation, with one participant contributing, “[My primary evaluator] allows personal differences to get in the way of his objective evaluation.” Participants were also concerned about the consistency of the evaluation process. One participant, who had multiple evaluators involved in the overall process, expressed, “The evaluations are not consistent from administrator to administrator.” All emergent themes, codes, example quotations, and response frequencies by theme can be found in Appendix G.

Finally, participants indicated how evaluative information was shared and reviewed following completion of their most recent teacher evaluation. A total of 274 participants (82%) specified they had a conference with their primary evaluator to review the evaluation, while 264 participants (79%) received a written evaluation summary. A total of 115 participants (34%) received brief verbal reports summarizing their performance, and six participants (2%) indicated they did not receive or expect to receive any feedback about their performance. When asked if participants received feedback in any other manner, a total of 11 respondents referred to online or electronic feedback. Others mentioned they received completed rubrics or checklists ( $n = 2$ ), while one participant mentioned receiving specific professional development assignments intended to address areas needing improvement.

### **Music Teachers' Perceptions of Teacher Evaluation Qualities, Benefits, and Impact**

A second purpose of this study was to examine K–12 music educators' perceptions about various aspects of the teacher evaluation process. More specifically, in this study I sought to determine the degree to which K–12 music educators' found teacher evaluations to be clear, fair, and useful. An additional purpose was to determine the level of impact teacher evaluation processes had upon K–12 music educators' daily practice.

**Evaluation qualities.** Seven Likert-type items (1 = strongly disagree, 6 = strongly agree) were used to elicit K–12 music teachers' overall impressions of the evaluation process, such as its clarity and value. Negatively worded items (e.g., “Our teacher evaluation process is not worth the effort”) were reverse-scored. While participants were, on average, of a relatively neutral position regarding the clarity of the teacher evaluation process ( $M = 3.29$ ,  $SD = 1.17$ ), more participants believed the process was clear (with  $n = 158$  either agreeing or strongly agreeing) than not (with  $n = 84$  either disagreeing or strongly disagreeing). Participants tended

to respond negatively ( $M < 3.00$ ) to all other items that addressed matters of accuracy, equity, and value. In particular, participants were not satisfied with the evaluation process ( $M = 2.41$ ,  $SD = 1.09$ ). Means and standard deviations for participants' overall impressions of the evaluation process are summarized in Table 4.08.

**Table 4.08**

*Means and Standard Deviations for Participants' Overall Impressions of the Evaluation Process (n = 294)*

<b>Evaluation Aspect</b>	<b>M</b>	<b>SD</b>
Our teacher evaluation process is clear to me.	3.29	1.17
I would choose to participate in our teacher evaluation process even if it was not required.	2.90	1.20
Our teacher evaluation process is not worth the effort.*	2.83	1.15
Our evaluation process provides an accurate depiction of my teaching performance.	2.53	1.13
Our teacher evaluation process equitably assess my performance.	2.51	1.09
My school should continue to implement the current evaluation process.	2.45	1.08
I am satisfied with our teacher evaluation process.	2.41	1.09

*Note.* 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree. \*Item reverse scored.

A total evaluation impressions subscale score was calculated by summing responses for each of the evaluation impression items. The reliability estimate for this subscale was strong

( $\alpha = .89$ ). The mean total evaluation impression score was 18.92 ( $SD = 6.18$ , possible range: 7–35), which indicates that, on average, participants were ambivalent about important qualities of the evaluation process. However, the large standard deviations suggest that music teachers' views may not be uniform.

**Evaluation benefits.** Participants reported their impressions of evaluation benefits by responding to six Likert-type scale items that used the 6-point response format describe above in relation to evaluation process qualities. Items described potential benefits, such as improved student learning and an increase in reflective practice. On average, participants indicated slight disagreement with all of the statements describing potential benefits ( $M < 3.00$ ). The item regarding reflective teaching practice (“The evaluation process causes me to reflect more upon my teaching”) had the most favorable response ( $M = 2.99$ ,  $SD = 1.21$ ). Table 4.09 summarizes means and standard deviations for participants' overall impressions of the potential benefits of teacher evaluation.

**Table 4.09**

*Means and Standard Deviations for Participants' Overall Impressions of Evaluation Benefits (n= 294)*

<b>Evaluation Benefit</b>	<b>M</b>	<b>SD</b>
The evaluation process causes me to reflect more upon my teaching.	2.99	1.21
The evaluation process motivates me to improve my teaching practice.	2.71	1.19
I have used feedback from the evaluation process to help identify relevant professional development opportunities.	2.69	1.16
The evaluation process has helped me become a more effective teacher.	2.62	1.10
The evaluation process contributes to improved student learning.	2.52	1.09
The achievement of my students has improved as a result of our evaluation process.	2.33	1.02

*Note.* 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly Agree.

A composite evaluation benefits score was calculated by summing participants' responses across the six evaluation benefits items included in this subscale. The reliability estimate for the benefits subscale was strong ( $\alpha = .92$ ). The mean composite evaluation benefits score was 15.86, with a standard deviation of 5.76 (possible range: 6–36). Overall, these findings suggest participants did not view the evaluation process as being particularly beneficial to them as professional music educators.

**Evaluation impact.** Finally, participants were asked to indicate the degree to which they

believed the teacher evaluation process impacted daily teaching practices. Music teachers reported their impact perceptions by responding to six items that used the 5-point, Likert-type format described previously. In general, participants reported neutral to negative opinions when considering the impact teacher evaluation had on their daily practice as K–12 music teachers. Participants perceived teacher evaluation to have the least positive impact on their ability to select appropriate instructional materials ( $M = 2.82$ ,  $SD = 0.88$ ). Descriptive results for the impact items are displayed in Table 4.10.

**Table 4.10**

*Means and Standard Deviations for Participants' Overall Impressions of Evaluation Impact (n = 293)*

<b>Area of Impact</b>	<b><i>M</i></b>	<b><i>SD</i></b>
My ability to implement appropriate instructional methodologies.	3.10	0.76
My ability to utilized planning and preparation time in a manner that best suits students.	3.03	0.73
My ability to make independent choices regarding my daily classroom practice.	2.96	0.81
My ability to assess students in a manner relevant to my content area.	2.91	0.93
My ability to utilize instructional time in a manner that best benefits students.	2.89	0.90
My ability to select appropriate instructional materials.	2.82	0.88

*Note.* 1 = Extremely negative impact, 2 = Negative impact, 3 = No impact, 4 = Positive impact, 5 = Extremely positive impact.

A composite evaluation impact score was calculated by summing participants' responses across the six evaluation items. These composite scores proved highly reliable ( $\alpha = .94$ ). The mean composite evaluation impact score was 17.69, with a standard deviation of 4.39 (possible range: 6–30). Overall, these findings suggest that participants felt relatively neutral about the impact teacher evaluation had on their daily practice.

### **Relationships Among Teacher Perception Subscales and Teaching Experience**

As described earlier, the questionnaire explored four facets of music teachers' perceptions regarding the evaluation process. These four facets included (a) the evaluators' approach and qualifications, (b), qualities of the evaluation process, (c) benefits of the evaluation process, and (d) the potential impact of the evaluation process on daily classroom practice. Relationships among composite subscale scores, representing each evaluation facet were examined using Pearson correlations. Total years of teaching experience were also considered in the Pearson correlations, as this variable is commonly examined in relationship to teachers' attitudes in times of educational change. Assumptions for linearity and normal distribution were checked and satisfied (Morgan, Leech, Gloeckner, & Barrett, 2011).

Music teachers' perceptions across all four evaluative dimensions (characteristics, evaluator qualities, benefits, daily instructional impact) were positively correlated at a statistically significant level ( $p < .01$ ), and all relationships were of a modest to moderately strong magnitude (ranging from  $r = .34$  to  $r = .68$ ). This provides evidence that the four facets of the teacher perceptions were related, but conceptually distinct. There were no significant relationships, however, between teaching experience and any of the teacher perception subscales. Scatterplots confirmed all non-significant correlations were not due to curvilinear relationships. The correlation matrix appears in Table 4.11.

**Table 4.11**

*Pearson Correlation Matrix for Facets of Music Educators' Perceptions of Teacher Evaluation and Total Years Teaching Experience (n = 288)*

	<b>Overall Impressions</b>	<b>Benefits</b>	<b>Daily Impact</b>	<b>Years Teaching</b>
Perceptions of Evaluator	.51**	.41**	.34**	.09
Overall Impressions		.68**	.60**	.07
Benefits			.64**	.03
Daily Impact				-.01

*Note.* \*\*Correlation is significant at  $p < .01$  (2-tailed).

### **Analyses of the Effects of Individual Differences on Teacher Perceptions**

To investigate the causal-comparative effect of individual group differences upon music teacher perceptions of the evaluation process (e.g., primary evaluator approach and competency, evaluation process qualities, evaluation benefits, evaluation impact upon daily practice), a series of multivariate analyses of variance (MANOVA) were conducted. Individual difference variables included gender, tenure status, teaching level (e.g., K-5, 6-8, 9-12, mixed), and teaching area (e.g., general music, instrumental music, choir). Assumptions for sampling and independent observations, homogeneity of variance (Levene's test), homoscedasticity (Box's M test), and multicollinearity were checked and met. The assumption for multivariate normality was violated; however, MANOVA is generally robust to violations of multivariate normality, so results can be interpreted with confidence.

Among the grouping variables, significant multivariate effects were evident for gender ( $\lambda = .944, p = .040$ ), area ( $\lambda = .899, p < .001$ ) and level ( $\lambda = .889, p = .001$ ). The effect size for each of these variables, however, was small (Cohen, 1988), explaining less than 10% of the total variance. There were no statistically significant effects for tenure status. Table 4.12 summarizes the multivariate tests.

**Table 4.12**

*MANOVA Summary Table for the Effect of Individual Differences upon Teacher Perceptions (n = 288)*

<b>Variable</b>	<b>Wilks Lambda</b>	<b>F</b>	<b>df</b>	<b>p</b>	<b>Partial <math>\eta^2</math></b>
Gender	.944	2.040	8	.040	.028
Area	.899	3.802	8	< .001	.052
Level	.889	2.808	12	.001	.038
Tenure Status	.995	0.347	4	.846	.005

To determine which facets of music educators' teacher evaluation perceptions were impacted by gender, area, and level, follow-up univariate analyses of variance were conducted (ANOVA). Results indicated that gender ( $p < .05$ ), teaching area ( $p < .05$ ), and teaching level ( $p < .01$ ) all had a significant effect on music teacher perceptions of evaluation benefits ( $p < .01$ ). However, no other independent variables had a significant effect on any of the other dependent variables.

ANOVA results are summarized in Table 4.13. In addition to degrees of freedom, F values, and  $p$  values, effect size is reported as partial eta squared. According to Salkind (2010),

“Eta-squared quantifies the percentage of variance in the dependent variable (Y) that is explained by one or more independent variables (X)” (p. 423). All effect sizes determined in follow-up univariate ANOVA analyses were small, as no independent variable accounted for more than 6% of the variance in any of the dependent variables.

**Table 4.13**

*Follow-up ANOVA Summary Table for Individual Differences on Teacher Perceptions (n = 288)*

<b>Dependent Variable</b>	<b>df</b>	<b>F</b>	<b>p</b>	<b>Partial <math>\eta^2</math></b>
<b>Perceptions of Evaluation Process</b>				
Gender	2	1.594	.205	.011
Area	2	0.435	.648	.003
Level	3	0.401	.752	.004
<b>Perceptions of Evaluation Impact upon Daily Practice</b>				
Gender	2	1.448	.237	.010
Area	2	0.519	.596	.004
Level	3	1.353	.257	.014
<b>Perceptions of Evaluation Benefits</b>				
Gender	2	3.897	.021*	.027
Area	2	5.377	.005**	.037
Level	3	5.931	.001**	.059
<b>Perceptions of Primary Evaluator</b>				
Gender	2	.884	.414	.006
Area	2	1.874	.155	.013
Level	3	2.019	.111	.021

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

Scheffe post-hoc tests were conducted to determine where group differences for teaching level and teaching area existed in relation to perceived evaluation benefits. Results indicated that elementary school music educators (grades K–5,  $M = 17.79$ ,  $SD = 5.58$ ) demonstrated more positive perceptions of the benefits of teacher evaluation than did high school music educators (grades 9-12,  $M = 14.27$ ,  $SD = 5.35$ ) ( $p = .001$ ). Similarly, general music teachers ( $M = 17.46$ ,  $SD = 5.93$ ) demonstrated more positive perceptions of teacher evaluation's overall benefits than did instrumental music teachers ( $M = 14.99$ ,  $SD = 5.57$ ) ( $p = .006$ ). This result is not surprising, as the majority of elementary teacher participants (98.6%,  $n = 72$ ) also taught general music. Females ( $M = 16.75$ ,  $SD = 5.38$ ) demonstrated more positive perceptions of the benefits of teacher evaluation than did male music educators ( $M = 14.86$ ,  $SD = 5.95$ ) ( $p = .040$ ). Means and standard deviations for perceptions of evaluation benefits as revealed by each independent variable are summarized in Table 4.14.

**Table 4.14**

*Scheffe Post-Hoc Results for Perceptions of Evaluation Benefits (n = 288)*

<b>Independent Variable</b>	<b>M</b>	<b>SD</b>	<b>N</b>
Area			
General Music	17.47 <sup>a</sup>	5.96	88
Vocal	15.48 <sup>ab</sup>	5.53	54
Instrumental	14.99 <sup>b</sup>	5.57	143
Level			
K-5	17.79 <sup>a</sup>	5.59	73
Mixed	17.31 <sup>ab</sup>	7.83	16
6-8	15.80 <sup>ab</sup>	5.50	101
9-12	14.27 <sup>b</sup>	5.35	98

*Note.* Scheffe tests of mean differences (a versus b) were statistically significant.

## Section B: Open-Ended Items

### Open-Ended Data Collection and Preparation for Analysis

In addition to the 14 closed-ended items, two optional open-ended items were included to provide more personalized and nuanced information regarding participants' experiences with and perceptions of the teacher evaluation process. The first open-ended item, "Describe how the teacher evaluation process at your school impacts music student learning and/or your effectiveness as a music teacher," asked participants to reflect upon the practical implications of the evaluative process. The second open-ended item read: "How has the teacher evaluation process impacted your perception of music teaching as a *profession*? Does this process elevate or diminish the professional status of music teachers?" This question focused upon the broader, professional implications of the evaluative process. In using the Qualtrics survey platform to develop the questionnaire, I required participants to complete all closed-ended items, but not the two open-ended items. The rationale for this decision is that open-ended items require more response time and/or may be viewed as invasive, such that requiring completion could reduce the response rate.

All participant responses to open-ended items were recorded and stored in the Qualtrics online survey platform. Upon the close of the data collection window, responses for each open-ended item were exported to Microsoft Excel 2011 version 14.0.0 for Mac, and data were subsequently transferred to Microsoft Word 2011 version 14.0.0 for Mac. Responses were then organized into a table that included the participant number, the response itself, and a column allowing for code labels for each individual response. Separate files were kept for each of the two narrative-response open-ended items. Once the initial data files were established, two

additional files were created in Microsoft Word. These files served as the codebooks for each of the two major open-ended questions.

### **Coding Open-Ended Responses**

According to Miles and Huberman (1994), “*codes* are tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study” (p. 56). The authors emphasize that when examining inferential information, codes help highlight the meaning of the words themselves. Coding also helps the researcher synthesize information into manageable, organized patterns. Prior to coding open-ended responses, I first read through all narrative data to gain a general sense of participants’ responses (Creswell, 2009). Next, I used an inductive coding approach (LeCompte & Schensul, 1999), developing both descriptive and in-vivo codes (Saldana, 2012). To help manage the codes, I assigned each code a color label and highlighted associated text with the appropriate color. After coding open-ended responses, codes were organized into patterns that reflected broader themes, including “impact,” “concerns,” and “teacher reaction.” Pattern coding also revealed positive, negative, and neutral valence, particularly with regard to “impact” (Miles & Huberman, 1994). Finally, I created a preliminary codebook organizing broader patterns, associated subcodes, and two representative examples illustrating each code. Frequency counts were also calculated and incorporated into the preliminary codebook to ascertain the prevalence of each emergent pattern.

**Verification strategies.** To establish the trustworthiness of code assignments and pattern-level analysis for the open-ended items, external audits were conducted with experts in qualitative data analysis. Once all responses for the first and second open-ended items were coded, preliminary codebooks were developed. A random sample of 15 responses with a minimum of 30 total possible codes was then sent to the first expert external auditor along with

the codebooks to determine inter-coder agreement on all assigned codes. According to Miles and Huberman (1994), inter-coder agreement of 70% is acceptable, though inter-coder agreement of 80% or higher is preferable. Initial inter-coder agreement was satisfactory (75%).

Following the initial external audit, email communication involving check-coding was employed to sharpen codes and increase reliability (Miles & Huberman, 1994). The first external auditor highlighted several opportunities for code clarifications, including adding a code that encompassed participants' concerns with evaluative criteria, rather than the initial all-encompassing code that addressed concerns with the evaluation process as a whole. After reviewing participant responses, the need for a code regarding evaluative criteria was substantiated, and a code with associated representative examples was added to the second iteration of the codebook. Additional email communication with the external auditor focused on coding disagreements, particularly disagreements that involved inferences as opposed to explicit statements. This issue was apparent in the following response provided by a female high school band director in Florida:

The teacher evaluation process has nothing to do with music. It was designed 35 years ago and involves a math formula that nobody can figure out – but here's your score.

In this instance, the external auditor initially coded the response assuming the participant implied a desire for a more effective evaluation tool. However, I coded this response to reflect the participant's perceptions on the evaluation system's lack of relevance and fit to music teaching. After conferring with the external auditor, we agreed to code similar items at face value; in other words, codes were assigned as per participants' explicit statements.

In addition to email communication, I had an in-person meeting with the first external auditor to further discuss and refine codes. In this meeting, we revised the codebook after carefully examining response examples that arguably fit into more than one code category. We also identified larger patterns that emerged from coded responses, in order to bring a more broad and dynamic perspective to the interpretation of open-ended data. Once a revised codebook was established, a second external auditor was provided a random sample of participant responses along with the new codes. Inter-rater agreement remained at 75%.

The second open-ended item targeted music educators' sense of professionalism within the context of teacher evaluation reform. More specifically, participants were prompted to consider how the teacher evaluation process impacted their perception of music teaching as a *profession*, and whether the process elevated or diminished the professional status of music teachers. To ensure data were again reliably analyzed, an external audit was conducted with an expert in open-ended coding strategies. I coded responses for the second open-ended item and created an additional codebook. I then sent a random sample of 15 responses with a minimum of 30 total possible codes to the expert external auditor along with the codebook to determine inter-rater agreement on all assigned codes. Initial inter-coder agreement was slightly lower (71%) than agreement for the first open-ended item.

Upon finishing the first round of coding, it became evident that more condensed codes might capture broader conceptualizations of the emergent themes for the second open-ended item. Details regarding the formulation of the condensed codebook can be found in Appendix H. I then recoded all data, and compared the new codes against those assigned by the external auditor. Several of the responses originally yielding coding disagreements now fell uniformly into the new, broader codes. A final iteration of the codebook was established, and a subsequent

round of external auditing took place. To ensure trustworthiness, I provided a separate external auditor with a random sample of participant responses along with the revised codebook. The resulting inter-coder agreement level was more satisfactory (75%). All tables presented in this section represent the final iterations of codes and code definitions for both open-ended items.

Codes for both open-ended items reflected three major themes: (a) impact, (b) concerns, and (c) teacher reactions. *Impact* refers to the various ways in which teacher evaluation affected music teachers, both in general, and specific to their professional practice, student learning, and professional status. *Concerns* refers to music teachers' apprehension about various aspects of the teacher evaluation process, including the qualifications of the evaluator, time demands, and the overall fit of the evaluative criteria. *Teacher reactions* refers to the ways in which music teachers responded to the demands of the evaluation process, either by "playing the game" to meet certain professional expectations, or through relying upon their own creativity and initiative to ensure the quality of their practice was maintained or developed throughout the evaluation process.

### **Open-Ended Item #1**

The first open-ended item read: "Describe how the teacher evaluation process at your school impacts music student learning and/or your effectiveness as a music teacher." Of the 293 participants that reached the open-ended items, a total of 246 participants (84%) opted to respond to the first item. Because participant responses typically touched upon multiple themes, a total of 445 statements were coded for the first open-ended item.

**Theme: Impact.** Impact-related codes and their related response frequencies are summarized in Table 4.15. These response frequencies indicate the number of participants that made remarks that fit within the parameters of each code.

**Table 4.15**

*Codes for Responses Reflecting “Impact” Theme (n = 246)*

	<b>Code</b>	<b>Theme</b>	<b>Example 1</b>	<b>Example 2</b>	<b>Frequency</b>	<b>% of Item #1 Total Respondents</b>
<b>No Impact</b>	NiGE	No Impact in General	No impact.	I think the process does not impact music education much at all.	48	20%
	NiSL	No Impact on Music Student Learning	The teacher evaluation process does not directly impact music student learning at my school.	The state evaluation tool has no bearing on the success of my students.	30	12%
	NiTE	No Impact on Teaching	The current evaluation process at my school has no impact on my teaching.	I don't do anything different in my instruction than I normally would.	39	16%
<b>Negative Impact</b>	NeGE	Negative Impact in General	It is a negative distraction.	Negative impact.	3	1%
	NeSL	Negative Impact on Music Student Learning	The evaluation process actively hinders music student learning.	The evaluation process slows down student learning in the music classroom.	12	5%
	NeTE	Negative Impact on Teaching	It slows down the natural flow of teaching.	The LEQ'S that we are required to post and use do limit teachable moments. They also limit my ability to help classroom teachers by reinforcing their current content.	21	9%
<b>Positive Impact</b>	PoTE	Positive Impact on Teaching	Helps keeps me on task, gives me direction and objectives for growth.	I have spent more time reflecting on my teaching strategies. I have also endeavored to seek out more development opportunities for myself.	37	15%
	PoSL	Positive Impact on Music Student Learning	The reflection on teaching practice has brought about new approaches and prioritization that has had a positive impact on student learning.	The process has a very clear rubric for teachers, and provides specific examples of behaviors and patterns which lead to higher student achievement	5	2%

**No impact.** A total of 48 (20%) participants indicated the teacher evaluation system had no impact in general (NiGE). Responses included, “It does not have any affect,” or, “It doesn’t affect it either way.” Because such responses lacked specificity with respect to teaching effectiveness or student learning, these responses were coded as “no impact in general” (NiGE).

Other participants were more specific in their responses and observed that the teacher evaluation process had no direct impact on professional practice and/or teaching effectiveness (NiTE). For example, a teacher in Tennessee remarked, “The current evaluation process at my school has no impact on my teaching,” while another teacher commented, “I don’t do anything different in my instruction than I normally would.” Some participants alluded to a neutral or negligible impact. One teacher pointed out, “Nothing was gained from my evaluation to make me a better educator,” while another wrote, “As of right now, [the teacher evaluation process] does not make me more or less effective.” A total of 39 participants (16%) viewed the teacher evaluation process as having little or no impact on teaching practice and teaching effectiveness.

In terms of student learning, a total of 30 participants (12%) believed that the teacher evaluation process had no impact (NiSL). According to one participant, “I don’t believe the evaluation process that we use impacts the students’ learning,” while another participant echoed, “The evaluation process really has little or no impact on student learning.” Overall, comments alluding to the teacher evaluation process having no substantive impact on teaching effectiveness and student learning, accounted for over a quarter of all coded responses ( $n = 117, 26\%$ ). For additional examples of each “no impact” code, refer to Table 4.17.

**Negative impact.** Only 15% of comments ( $n = 36$ ) attributed negative impacts to the teacher evaluation process, be they in general (NiGE), in the context of their teaching (NiTE),

or as related to student learning (NiSL). One participant referred to the evaluation process as “strangling,” and another referred to teacher evaluations as “a negative distraction.” Among participants who felt the teacher evaluation process negatively impacted student learning, one participant wrote, “Our district’s chosen evaluation process is a serious interruption to my students’ learning.” Another contributed, “The students do not perform as well as they used to, in a musical way.” Approximately 9% of participants ( $n = 21$ ) referenced a direct negative impact upon their teaching practice, with one participant reflecting, “[The teacher evaluation process] slows down the natural flow of teaching,” and another writing, “I have to focus more on teaching vocabulary and less on teaching musical skills.”

**Positive impact.** Approximately 15% of music teachers ( $n = 37$ ) believed the teacher evaluation process had a positive effect on various facets of their practice, including instructional planning. For example, one participant explained, “My plans are much more closely aligned to National Standards.” Another participant reflected upon how clear evaluative criteria assisted with effective lesson planning:

One of the most important impacts on me has been preparing my lessons. I can better plan my lessons when I look at the standards and the evaluation rubric together. When I know what is expected of me and what is expected of my students, then I can prepare an effective lesson for all of us.

Other participants believed that teacher evaluations had positive impact upon the delivery and pacing aspects of their professional practice:

Receiving regular feedback from someone [outside of music] assists me in ensuring my lessons are clear, concise, and effective. I also receive tips on improving management, pacing, and structure, which help me as a teacher to make the most effective use of my time.

Several teachers observed how the evaluation process fostered meaningful reflection that enhanced their practice, while others found the evaluation process helped them to identify best teaching practices in their music classroom. Despite many teachers believing the evaluation process had a positive impact on their teaching, only a handful of participants ( $n = 5$ ) described the evaluation process as having a positive impact upon student learning, relating that the evaluative process helped contribute to “higher student achievement” and “increased student engagement.”

Overall, the largest proportion of teachers who responded to the first open-ended item provided comments that implied the evaluation process had no direct impact upon them, while the remaining participants were split, reflecting either positive or negative views on impact (primarily as related to teaching practice). In all instances, comments on how evaluation impacted student learning were more rare.

**Theme: Concerns.** Participants expressed a variety of concerns about the teacher evaluation process. These included concerns about the time demands associated with the evaluation process (CoTI) and concerns about the evaluator (CoEV). Additionally, participants expressed concerns about the overall fit of the evaluation process. More specifically, participants were concerned about the validity of growth data used in their evaluations (CoFITGro) and the overall limitations of the scope of the evaluative process (CoFITSco).

Representative examples and response frequencies for each of the concern categories (time, evaluator, and fit) are summarized in Table 4.16.

**Table 4.16**

*Codes for Responses Reflecting “Concerns” Theme (n = 246)*

	<b>Code</b>	<b>Theme</b>	<b>Example 1</b>	<b>Example 2</b>	<b>Frequency</b>	<b>% of Item #1 Total Respondents</b>
<b>Time</b>	CoTI	Concerns with Time	The teacher evaluation process demands time spent on irrelevant paperwork. It robs me of time to focus on music education needs.	I am overwhelmed with new evaluation implementation without adequate training time.	35	14%
	CoEV	Concerns with the Evaluator	Evaluators are not appropriately trained to see how music teaching and rehearsal techniques fit within the context of the rubric	Most evaluators know nothing about how music teachers teach.	40	16%
<b>Evaluator</b>	CoFITGro	Concerns with Fit – Validity of Growth Data	It is my understanding that a large part of my evaluation depends on the total scores from my school, which consist of scores on state tests from students I do not teach, given by teachers who do not know "beans" about teaching music.	I am sad that reading and math scores are a part of my evaluation as an elementary general music teacher.	15	6%
	CoFITSco	Concerns with Fit – Limited/ Unrealistic Scope	Our current evaluation process does not allow for the differences in teaching a music (or other performing arts) class as opposed to academic or core subjects	My school uses the same teacher evaluation process for all teachers and subject areas. This objectives assessed have little to do with music education or the music curriculum. Because of this, I feel that the evaluation is not effective or appropriate for arts teachers.	72	29%
<b>Fit</b>						

***Time concerns.*** Participants' perceptions of the time demands associated with teacher evaluation constitute *time concerns*. Around 14% of participants ( $n = 35$ ) cited *time* as being the predominant concern with regards to the evaluation process. As one participant described:

The biggest negative impact is the extreme loss of time. I have spent countless hours on iObservation. I spent 5 hours working with another music teacher to create our developmental growth plans. I spent 2 hours filling out a pre conference form for my formal observation. I spent another hour doing the post conference. I spent an hour typing up reflection logs and attaching them to iObservation. And of course, I have spent 1-2 hours per student learning objective (we have 4 in our developmental growth plans) charting data (4-8 hours for that total). There is also loss of instructional time. I have 45 minute periods, and I would estimate I lose 5-15 minutes per class period to do the required learning goals/rubrics/reflections, usually 5 minutes at the beginning and 5-10 minutes at the end to reflect/evaluate based on the rubric. I am not saying you should not have a goal or rubric, but the documenting is killing me.

Other participants referred to the evaluation process being a "waste of time," referring to "countless hours doing things to check off of a list." Participants did not make any positive comments regarding evaluation and time.

***Evaluator concerns.*** The majority of participants' concerns regarding the supervisory aspect of their evaluation process reflected trepidations about the primary evaluator's qualifications for assessing music teacher effectiveness. In particular, participants' described evaluators' lack of knowledge about music. As one participant noted, "Most evaluators know

nothing about how music teachers teach,” while another wrote, “I have never been formally evaluated by a single person that was trained in music.” Specifically regarding the classroom observation component of teacher evaluation, one participant described:

Evaluators are not appropriately trained to see how music teaching and rehearsal techniques fit within the context of the rubric and that the [evaluation rubric requirements] look oftentimes very different in performance-based classes.

A total of 40 participants (16%) expressed some level of concern with the evaluator(s) involved in the overall evaluation process.

***Fit concerns.*** The third and more broad theme, *fit concerns*, references participants’ perspectives as to the appropriateness and value of the evaluation process. *Fit concerns* included concerns with the validity of growth data used in music teacher evaluations (CoFITGro) as well as concerns about the applicability of the scope and expectations of the evaluative process (CoFITSco). A total of 15 participants (6%) described concerns about growth data (i.e., being evaluated on the basis of non-music test scores). As one participant wrote, “My biggest concern is that half of my evaluation score is derived from math and reading scores on the state assessment.” Another participant stated, “Being evaluated based on school district test scores instead of my student's learning seems unbalanced.”

Nearly one-third (29%,  $n = 72$ ) of all participants reported some level of trepidation as to the scope and applicability of evaluative expectations. According to one participant, “Our current evaluation process does not allow for the differences in teaching a music (or other performing arts) class as opposed to academic or core subjects.” Several participants

mentioned how the evaluation process lacked alignment with the unique demands of the music classroom, with one participant writing that the evaluation process had, “too many cookie-cutter aspects that do not fit well into a music curriculum.” Other participants referred to the evaluation process as “too generic,” arguing that “one size does definitely not fit all” with regard to evaluating music teaching.

**Theme: Teacher reaction.** A large number of responses to the first open-ended item depicted how participants reacted to the evaluation system. These responses were often manifested as increased self-reliance (RelSel) or resignation to “jump through hoops” (HoSho) in order to make the system work for them or to satisfy demands of the process. Other reactions included participants’ marked desire for a more appropriate evaluation tool for music teachers (DeTo). Reaction-related codes and associated frequencies are summarized in Table 4.17.

**Table 4.17***Codes for Responses Reflecting “Teacher Reaction” Theme (n = 246)*

<b>Code</b>	<b>Theme</b>	<b>Example 1</b>	<b>Example 2</b>	<b>Frequency</b>	<b>% of Item #1 Total Respondents</b>
HoSho	Hoop Jumping/Put on a Show	I feel like I have to put on a "show" to show I'm teaching in a certain way that doesn't fit my teaching personality.	The teacher evaluation process is a group of hoops through which I am expected to jump.	30	12
RelSel	Reliance upon Self in Evaluation Process	I do what I do for my kids because I require it of myself.	I do what I know is best.	24	10%
DeTo	Desire for More Appropriate Tool	In order to be effective for our discipline, a more arts-specific evaluation needs to be developed.	A more appropriate tool would, I believe rate me just as highly but would give me better feedback and allow me to improve more.	12	5%

***Jumping through hoops.*** A total of 30 participants (12%) described the teacher evaluation process as a “hoop” through which they had to jump in order to achieve a satisfactory overall evaluation, particularly in the realm of the classroom observation component. This theme reflected participants’ comments regarding temporary, conscious adaptations to their practice intended to satisfy evaluation expectations. One orchestra director explained, “I modify what I say and do to fit in with the evaluation.” One participant with twelve years of teaching experience offered, “Teachers are forced to choreograph a lesson to ensure that each expected behavior is seen.” The temporary adaptation of practice is highlighted by another participant, who explained, “Evaluations are isolated events where music teachers jump through hoops designed for core teachers, then return to teaching music when evaluations are over.” Another response, “I was told to ‘put on a show’ at my lesson in order to get all my boxes checked,” suggests that music teachers might be encouraged by other stakeholders to modify their behaviors for the evaluation process.

***Self-reliance.*** A total of 24 participants (10%) noted how they relied upon their own sense of priorities and values throughout the evaluation process. Sometimes this self-reliance reflected a business-as-usual approach, where teachers continued to do their “own thing” despite evaluative demands. One participant wrote, “I have always done my best without worrying about being evaluated.” Other participants described self-reliance in relation to how they improve their practice in spite of an evaluation system they find misaligned to music teachers’ needs. As one participant described, “I continue to improve as a music teacher because I seek out enrichment/staff development opportunities that I know will help me,” while another participant contributed, “My student's performance in class and at concerts & festivals has a greater impact on my desire to be a better teacher than the teacher evaluation process does.”

*Desire for more appropriate tool.* The perceived lack of alignment between the evaluation process and the unique demands of music teaching was articulated by a handful of participants ( $n = 12$ , 5%) who expressed a desire for a more effective, appropriate evaluation tool. As one participant wrote, “If our county adopted an “Arts Evaluation” system, which is used in some counties in my state, that would have a bigger influence on both students and my effectiveness.” Another participant expressed, “Until there is a standardized assessment for music the evaluation process will not be beneficial.” While many participants inferred a desire for a more appropriate evaluation system for music educators, only those responses that explicitly stated this desire were coded as “DeTo.”

**Section summary.** The first open-ended item was designed to solicit music teachers’ perspectives as to how the current teacher evaluation process impacted their perceived effectiveness and impact upon student learning. Responses varied in terms of specificity (discrete impacts identified in some instances and general critiques in others) as well as valence (negative, negligible, or positive impacts). While many participants expressed trepidation about the evaluation process, a sizeable subgroup of music teachers felt that the evaluation process did, in fact, have a positive impact upon their teaching practice. In fact, the number of coded responses linked to positive impacts upon teaching effectiveness was nearly twice that of responses associated with negative effects. However, participants also described a variety of concerns that suggested a level of dissatisfaction with the evaluation process. These concerns gave dimension to the perceived negative impacts upon music teaching and learning, which in turn informed participants’ reaction to the evaluation system.

### **Open-Ended Item #2**

The second open-ended item read: “How has the teacher evaluation process impacted

your perception of music teaching as a *profession*? Does this process elevate or diminish the professional status of music teachers?” The purpose of this question was to determine music teachers’ perceptions of the broader professional implications of the teacher evaluation process. With the exception of one individual, all participants who responded to the first open-ended item also responded to the second open-ended item. Additionally, four participants that did *not* respond to the first item chose to contribute to the second open-ended item. In total, 249 participants (85%) responded to the second open-ended item.

**Responses and major themes.** Final item codes for the second open-ended item reflected the same three primary themes: *impact*, *concerns*, and *reactions*. Participants often provided responses that offered a continuation to the first open-ended response. However, the sub-themes that emerged in the second open-ended item differed slightly from those of open-ended item number one. For example, reflecting the language in the second question, *impact*-related responses typically focused on *professional status*. The concept of *professional status* refers to the various ways in which teacher evaluation affected music teachers’ feelings of legitimacy and equality, particularly when compared with teachers in other academic disciplines. Responses falling under the category *concerns* still referenced issues of time, evaluator training and competence, and the overall fit of evaluation expectations in relation to music teaching and learning. However, these *concerns* also branched out to reflect participants’ worries about keeping their position. Finally, responses falling under the category *reactions* still addressed the effects of teacher evaluation upon participants’ self-reliance and need to “jump through hoops.” Additional sub-themes regarding the desire to leave the profession and thoughts about the evaluative process being an improvement upon past practice also surfaced.

**Theme: Impact on professional status.** Similar to responses to the first open-ended

item, participants reported that teacher evaluation had (a) no impact, (b) a negative impact, or (c) a positive impact upon music education as a profession. While some codes overlapped those present in the first open-ended item (e.g., PoTE, NeTE, NiGE), numerous unique codes emerged with regard to *impact*. Example quotations and response frequencies for codes supporting the theme *impact* are summarized in Table 4.18.

**Table 4.18**

*Codes for Responses Reflecting “Impact” Theme: Professional Status (n = 249)*

	Code	Theme	Example 1	Example 2	Frequency	% of Item #2 Total Respondents
<b>No Impact</b>	NiGE	No Impact on Perception of Profession	I don't think the evaluation process has changed my perception of music teaching as a profession.	I don't think the evaluation process itself can elevate or diminish the professional status of music teachers.	58	23%
	StaDimMu	Status of Music Educators/ Education Diminished	The process diminishes the professional status of music teaching – it's not as important as a grade level teacher.	If anything, the process makes my job seem unimportant, as if it were an afterthought.	92	37%
<b>Negative Impact</b>	StaDimEd	Status of Educators/ Education in General Diminished	[The process] diminishes the value of all teachers because the standards right now are set so high, nobody can reach them, yet when we don't reach them, we must not be a good teacher.	It has negatively impacted my perception of public education.	19	8%
	NeTe	Negative Impact on Teaching	I feel [the process] is an obstacle at this time to my daily teaching.	There are requirements and expectations that keep me from really teaching music.	8	3%
	MoraleLow	Negative Impact upon Morale	Teacher morale is the lowest I've ever seen in my 25+ years in the classroom.	Low morale is abundant as we are expected to help teach literacy (which I don't mind including) but when literacy instructions begins to trump my music instruction, I have a problem with that	20	8%
<b>Positive Impact</b>	StaElMus	Status of Music Educators/ Education Elevated	I think it elevates the status, as we are being rated based primarily on what we actually do in the classroom through the observations and the portfolio (collection of evidence).	The current policy of evaluating my delivery, and not my content, allows me to make professional decisions as a teacher. These decisions are part of professionalism, thus, the process elevates the professional status.	27	11%
	PoTE	Positive Impact on Teaching	The standards are high, and [the process] has really encouraged me to rethink the way I taught.	It has made me a better music teacher because it has truly caused me to be accountable for every minute.	9	4%

**No impact.** A total of 57 (23%) participants indicated the teacher evaluation process had *no* impact upon music educators' professional status. Participants approached this perceived lack of impact primarily from a personal perspective (e.g., "I don't think the evaluation process has changed *my* perception of music teaching as a profession" and "[Evaluation] has done nothing to help or harm *my* perception [of music teaching as a profession]")

***Negative impact.***

***Diminished status of music educators.*** Over one-third of participants (39%,  $n = 98$ ) indicated that, on some level, the teacher evaluation process diminished the professional status of music educators (StaDimMu). Participants connected beliefs about diminished status with feelings of inequality and comparative importance to other subject areas. One teacher wrote, "In my school, we are NRT (not real teachers)," while another wrote, "We are considered 'non-teachers' by some." Participants also reported general beliefs about music education "not mattering" to others. One participant reflected, "[Teacher evaluation] tacitly takes away our relevance," while another echoed, "[The teacher evaluation process] is simply a reflection of our education system's view of the arts – that they do not matter." Taking a similar position, an elementary music teacher in Tennessee contributed:

I feel it diminishes our status. We are being told, in a very subtle way, that we are no longer necessary, relevant, or wanted. Our subject is being marginalized and our students are being 'counseled' out of band into more 'rigorous' academic areas. This is a mistake!

Some participants also suggested that the "one size fits all" approach to teacher evaluation negatively impacted the professional status of music educators, expressing discontent

with regard to imposed conformity. A teacher in North Carolina expressed concern that the evaluation process was an attempt to “put all teaching into one box,” and another teacher from North Carolina referred to being “lumped in with everyone else.” Ultimately, there was an apparent dissatisfaction with the expectation for conformity to the expectations of other academic disciplines, with a resounding sentiment that such practices took away from the unique aspects of music teaching.

Other participants contributed feelings of worry, such as one participant that wrote, “The [music education] profession has been so damaged [by the teacher evaluation process], I do not know whether it will ever recover.” Largely, participants were not optimistic about the long-term impact of teacher evaluation upon the professional status of music education.

*Diminished status of all educators.* A total of 19 participants (8%) indicated that the teacher evaluation process negatively impacted the education profession as a *whole*, beyond the specific domain of music education (StaDimEd). As one participant wrote, “[The teacher evaluation] process has seriously eroded the professional status of all teachers,” while another contributed, “I believe [teacher evaluation] is a result of society blaming educators for everything they see going wrong in our country.” A handful of participants also explained that recent teacher evaluation policy has impacted their personal faith in the education system. One participant explained, “[Teacher evaluation] has negatively impacted *my* perception of public education.” Overall, a smaller percentage of participants expressed concern regarding teacher evaluation’s impact upon the professional status of teachers as a whole.

*Negative impact upon teaching.* A total of eight participants (3%) described the teacher evaluation process as negatively impacting their teaching, which in turn affected their sense of autonomy and overall professionalism. One participant wrote, “The teacher evaluation process

takes out the teacher's creativity and drive to make their classroom structure fit their individual style." Another participant described the teacher evaluation process as "taking away from the performance aspect of the class," while another described having to teach other subject areas within the music classroom in order to earn a satisfactory evaluation.

*Morale.* The concept of *morale* in the context of this study refers to teachers' feelings of professional confidence, purpose, and motivation. Approximately 8% ( $n = 20$ ) of participants described some form of discouragement or lowered morale as a result of teacher evaluation, with participants noting the "burdensome" nature of teacher evaluation having a "negative affect on attitude" (MoraleLow). One teacher wrote, "I think the current evaluation system is highly discouraging to many teachers," while a band director in Florida wrote:

This evaluation process makes my job more of a headache than one to be excited for each day. I try to stick with what is expected of me, but I find myself frustrated with the outcomes. I am able to achieve high ratings on my evaluations, but at what cost to my students' love for music? ...I feel I am trapped in the evaluation process.

These deflated feelings of morale align with the large percentage of music educators who also felt their professional status was diminished by the teacher evaluation process.

***Positive impact.***

*Elevated status of music educators.* A total of 18 participants (7%) believed the teacher evaluation process elevated their status as music educators (StaElMu). Some participants believed the increased demand in accountability contributed to elevated feelings of professionalism, with one participant stating that increased accountability "definitely legitimizes

our profession.” Supporting this notion, some participants described an appreciation for the ways in which teacher evaluation put them on “an equal plane” as teachers in other academic subject areas, effectively “leveling the playing field.” As one participant remarked, “It makes me [feel] like I am as valuable as a regular teacher.” Another stated, “We are now part of the rest of the faculty when it comes to standards.”

Other participants suggested the teacher evaluation system allowed for a degree of autonomy in their instruction, which empowered them as professionals. While some participants described how teacher evaluation elevated the status of music educators, no participants made mention of how teacher evaluation might elevate the status of teachers or the education profession as a whole.

*Improved teaching.* As in the first open-ended item, several participants ( $n = 9$ ) described the ways in which teacher evaluation improved their instructional practice, thereby elevating their personal professional standards and providing a rationale for amplified feelings of professionalism (PoTE). As one participant wrote, “[Teacher evaluation] has a strong impact on what I do in the classroom...[it] gives me direction on what I need to improve upon.” Another participant contributed that the high professional standards associated with the teacher evaluation process encouraged him to rethink how he taught, while a third participant stated, “[Teacher evaluation] encourages me to stay current...and not get boxed into a routine.”

**Theme: Concerns and professional status.** Similar to responses for the first open-ended item, participants expressed a variety of concerns regarding teacher evaluation and its impact upon music education’s status as a profession. These concerns again reflected themes surrounding the evaluator’s credibility, time demands, and overall relevance of the evaluation process to music teaching. Recurring *concerns* themes are summarized in Table 4.21.

In addition, several participants expressed unique concerns regarding the high-stakes consequences associated with teacher evaluation outcomes, including tenure and promotion status, as well as merit pay decisions. These reflections substantiated a new code, “Concerns over Status” (CoSTAT). Participants voiced apprehensions about the ways in which teacher evaluation impacted their professional status, in terms of tenure and promotion decisions as well as merit pay decisions. As one participant wrote, “If [other teachers] score high enough, they get upwards of \$15,000 extra dollars, while I get \$400 because I teach music. Talk about becoming disheartened and feeling unwanted.” One participant described the merit pay system as “creating a sense of competitiveness” among teachers, pitting music educators against traditional classroom teachers in a manner that denigrated the status of music education. Another participant described concern over losing her position all together, due to her evaluation being tied to student scores on tests outside of the music discipline. As teacher evaluation scores are commonly linked to tenure decisions, one participant wrote, “[Teacher evaluation] has greatly decreased my chances of getting tenure,” thereby complicating feelings about music education as a respected profession.

**Table 4.21**

*Codes for Responses Reflecting “Concerns” Theme: Professional Status (n = 249)*

	Code	Theme	Example 1	Example 2	Frequency	% of Item #2 Total Respondents
<b>Recurring Codes Also in Open-Ended Item #1</b>	CoTI	Concerns with Time	The amount of work and time spent on meeting the teacher evaluation requirements is ridiculous, and the powers that be simply have no idea.	In a nutshell, the data required to fulfill the evaluation, which is over and above the data we already collect in order to grade students, has become all-consuming.	17	7%
	CoEV	Concerns with the Evaluator	Evaluators who have no experience in a performance ensemble don't understand what's happening when they observe a class and can't effectively critique what they've witnessed.	It is difficult to have a non-music teacher help give concrete ideas about how a music teacher can improve his/her instruction.	31	12%
	CoFITSco	Concerns with Fit – Limited/ Unrealistic Scope	This process is not the most effective way to evaluate music teachers	The concepts and benchmarks that music teachers assess are not covered in the teacher evaluation process.	41	16%
	CoFITGro	Concerns with Fit – Validity of Growth Data	I am assessed based on the student achievement of students that I do not even teach.	My teaching evaluation score is based in part on how the school does overall on our state testing.	26	10%
<b>Unique Code to Open-Ended Item #2</b>	CoSTAT	Concern over Status	It has greatly decreased my chances of getting tenure.	I am concerned for my job. It could cost me my license, only time will tell.	8	3%

**Theme: Reactions and professional status.** The final theme in participant responses to the second open-ended item again reflected music educators' *reactions* to circumstances resulting from the teacher evaluation process, particularly as related to participants' perceptions of teacher evaluation's impact upon the professional status of music educators. These reactions still included the ways in which participants adapted their practice to meet prescribed expectations (HoSho), as well as the ways in which participants took independent initiatives to maintain their professional integrity throughout the evaluation process (RelSel). Reactions unique to the second open-ended item included participants' thoughts about remaining in the profession or not, due to the demands of the teacher evaluation process and its perceived lack of relevance to music teaching (NoTeach). Additionally, a small number of participants reflected that the current evaluation process was an improvement upon past practice (BeBo). Reaction codes, examples, and associated frequencies are summarized in Table 4.22.

**Table 4.22**

*Codes for Responses Reflecting “Teacher Reaction” Theme: Professional Status (n = 249)*

	<b>Code</b>	<b>Them</b>	<b>Example 1</b>	<b>Example 2</b>	<b>Frequency</b>	<b>% of Item #2 Total Respondents</b>
<b>Recurring Codes Also in Open-Ended Item #1</b>	RelSel	Reliance Upon Self in Evaluation Process	My own personal convictions give me professional status.	I find that I learn more about teaching and improving my classroom instruction by watching clinicians, bringing clinicians in, and discussing strengths and weaknesses with them than I do in my teacher evaluations.	17	7%
	HoSho	Hoop Jumping/Put on a Show	It is a dog and pony show for educators.	My arts colleagues and I see [teacher evaluation] simply as a hoop that we must jump through each year -- mandatory paperwork.	12	5%
	DeTo	Desire for More Appropriate Tool	Until there is an evaluation that is geared specifically toward music educators, there is no way that any evaluation can elevate the professional status of music educators	I think that the evaluation process has to be meaningful and comprehensive for it to actually motivate any of us to take it seriously.	8	3%
<b>Unique Codes to Open-Ended Item #2</b>	NoTeach	Desire to Leave Education/ Discourage Others	Honestly, [evaluation] is one of several factors I feel that are spurring me to change professions.	I no longer encourage my high school students to go into education. I almost tell them to run the opposite way! Something is wrong when you can't encourage someone to go into your own profession.	12	5%
	BeBo	Current Process Better Than Before	[The new system] is better than the “pass/fail” system that was used, but it is very lacking in specific content areas.	Overall, the evaluation system has been an improvement over the capricious evaluations of certain administrators.	3	1%

***Leaving the profession.*** A total of 12 participants (5%) reported a desire to either leave the profession of music teaching or an obligation to discourage others from pursuing a career in music education, while others expressed regret about choosing the teaching profession in the first place. One participant reflected:

[The teacher evaluation process] makes [you tell all to choose another profession... Good teachers are leaving. Folks that could be a teacher are jumping ship early in their career and are not giving education a chance.

Another participant reported already knowing several teachers that had left the profession *primarily because of the teacher evaluation process*, while another contributed, “I am going to leave the profession after 20 years, and I wouldn’t recommend it to anyone at this point.” These powerful statements reflect music educators’ diminished feelings of professionalism.

***Improved process.*** A handful of participants (1%,  $n = 3$ ) relayed that the current evaluation process was better than the system that was formerly in place (BeBo). As one participant commented:

Overall, the evaluation system has been an improvement over the capricious evaluations of certain administrators in that they are required to follow a certain rubric, making it more professional and less personal.

**Section summary.** The second open-ended item was designed to gather participants’ perceptions on how the teacher evaluation process impacted the status of music education as a

profession. Over one third of participants felt that the status of music education had been diminished in some capacity due to teacher evaluation, referring to feelings of lowered professional morale. A variety of other concerns suggested perceptions of diminished status, including concerns about the overall fit of the evaluative process, and how the scope of that process often failed to address the unique needs and contributions of music educators.

Participants' reactions to these circumstances ranged from "grinning and bearing it" to leaving the profession all together. Still, some participants did not perceive teacher evaluation to have any impact on music education's status as a profession. Implications for these findings will be discussed in Chapter 5.

## Chapter V

### Summary and Conclusions

#### Summary

The purpose of this study was to examine music teachers' perceptions of current teacher evaluation practices designed to comply with Race to the Top (RttT) initiatives being implemented in four states. In particular, I sought to explore key elements of the evaluation process, including who was responsible for conducting music teacher evaluations, which measures were considered in those evaluations, and what greater functions or objectives were to be achieved through the evaluative process. Moreover, in this study, I focused upon characteristics of the evaluation process, including its clarity, utility, and impact. Based on my review of related literature, it appears that no other researcher has examined the experiences of music educators subject to RttT teacher evaluation policies. By conducting a multi-state survey, a variety of approaches to teacher evaluation policy implementation could be considered.

In the spring of 2014, data were collected from 288 K–12 music educators in four states that had been awarded RttT grant funding, including Florida, North Carolina, Rhode Island, and Tennessee. Participation was delimited to those music educators who were members of their state Music Educators Association. The margin of error was  $\pm 5.65\%$  with a 95% confidence level. Participants completed the 23-item online Music Teacher Evaluation Index (MTEI). This survey instrument included 12 closed-ended and two open-ended items that addressed various dimensions of teacher evaluation, as well as nine items soliciting demographic information.

In this chapter, I begin by providing a brief summary of the major findings. I then describe the teacher evaluation process as perceived and experienced by

K–12 music educators in four RttT states (Research Question 1). Next, I explore music educators' perceptions of the fairness, clarity, and utility of teacher evaluation (Research Question 2), supporting this evidence with participants' responses to open-ended items. I then examine how teacher evaluation has impacted music classroom practice, including notions of professional growth and student achievement (Research Question 3). Following these discussions, I present participants' perceptions regarding the ways in which teacher evaluation has impacted music education as a profession (Research Question 4), connecting these perspectives with related research. After discussing results for each research question, I explore potential implications for both teacher evaluation policy and music educators. Finally, I discuss study limitations specific to sampling, methodology and data analysis before proposing potential directions for future research in the area of music teacher evaluation. Below are the aggregate results summarized according to each research question:

*What elements characterize the process used to evaluate music educators?*

- Participants report that RttT-driven teacher evaluation processes are more formative than summative in nature, despite high-stakes incentives and sanctions that are directly associated with evaluative outcomes.
- Music teachers lack clarity regarding whether or not evaluations are used to inform personnel decisions, even in states where evaluations are already tied to such decisions (e.g., Tennessee, Florida, Rhode Island). Similarly, music teachers lack clarity regarding whether or not student growth factors into their overall evaluation.

- Music educators report that classroom observations (both formal and informal) *are* and *should be* considered to a greater extent than other measures of teacher performance in the evaluation process.
- Participants moderately support other measures of teacher performance, suggesting that portfolios, professional development activities, and student growth data *should be* part of the evaluation process.
- Students' non-music standardized test score data represents the most common measure of student growth used in music teacher evaluations.
- Building administrators with no music expertise are the primary authorities in the evaluation process.

*How do music educators perceive the teacher evaluation process?*

- Most music educators are slightly dissatisfied with the teacher evaluation process and would prefer not to participate in the process if given the choice.
- A slight majority of participants believe they are clear on evaluation expectations.
- Music educators generally have good interpersonal relationships with their primary evaluator, but are ambivalent about whether or not that evaluator is qualified to accurately assess music teaching.
- Beyond concerns with the primary evaluator, open-ended responses indicate that music educators have reservations about the fit of the evaluation process, particularly given the unique nature of the music classroom. The time demands associated with teacher evaluations also trouble many participants.

*In what ways do music teachers believe evaluation practices impact and/or benefit the classroom?*

- Music educators report that teacher evaluation requirements have a slightly negative effect on classroom practices and routines.
- Some music educators acknowledge that the teacher evaluation process enhances professional self-reflection.
- Generally, music educators do not believe the evaluation process contributes to improved student achievement.
- Females find teacher evaluations to be more beneficial than do males, and general music teachers demonstrate more positive perceptions of evaluation's benefits than do instrumental music teachers. Elementary school teachers feel more positive about evaluative benefits than high school music teachers.
- Unintended consequences of teacher evaluation include the promotion of inauthentic teaching practices. Some participants report that evaluation expectations encourage classroom practices that do not align with traditional music classroom pedagogy.

*How do teacher evaluation practices impact music educators' perceptions of music teaching as a profession?*

- There is a lack of consensus as to whether teacher evaluation elevates or diminishes the status of music education.
- While some participants feel the evaluation process elevates music educators by holding them accountable to the same standards as other content area teachers, a larger number of participants believe the process negatively impacts music education by forcing music

educators to conform to a model ill-suited for their discipline.

- A handful of participants feel the evaluation process is detrimental to morale.
- Around one-quarter of participants feel the evaluation process has no impact at all on the professional status of music educators.
- Some music educators also express concerns regarding teacher evaluation's impact upon not just music education, but the teaching profession at large.

Overall, results suggest that music educators are somewhat unclear about and slightly dissatisfied with their evaluation process. In particular, music teachers are concerned about their evaluator's expertise, the type of student growth data considered in their evaluations, and the overall fairness and applicability of the evaluation process. Because these findings are aggregate representations of participant responses across each of four states, it is important to exercise discretion in interpreting findings.

## **Discussion**

### **Elements characterizing music teacher evaluation.**

*Evaluation purpose.* Participants indicated the extent to which common appraisal objectives were characteristic of their evaluation process. Some of these objectives were formative in nature ("to guide improvement of teachers' skills"), while others ("to make tenure and promotion decisions") were summative. Overall, participants reported that formative objectives were emphasized over summative objectives in their most recent evaluation process, corroborating research conducted by the National Center for Education Statistics (NCES) in 1994. In other words, despite the emphasis on outcomes, incentives, and sanctions within RttT teacher evaluation practices (Ravitch's 2010 concept of "punitive accountability"), music

educators in this study still perceived the evaluation process as largely designed to function as professional development. This finding suggests that Firestone's (2014) conception of intrinsically oriented professional motivations, which better reflect the "professional conception" of teaching described by Wise and Darling Hammond (1985, p. 30), may still be favored by building administrators and carry more weight with music teachers than punitive-oriented motivations. Administrators may also be hesitant to enforce sanctions among their teachers, motivating building leaders to instead emphasize formative evaluation objectives.

Embedded within music educators' responses, however, was a degree of uncertainty regarding summative functions. Even excluding participants from North Carolina (for whom evaluations will not inform personnel decisions until the 2016–2017 school year), roughly one-third of participants were unsure if their evaluations influenced tenure and promotion decisions, while another third of music teachers did not know if evaluations were intended to discharge incompetent teachers. As part of the RttT "Great Teachers and Leaders" initiative, teacher evaluation outcomes are intended to "be used as basis for policy and personnel decisions designed to improve student and school performance" (United States Department of Education, 2011, p. 24). Music teachers in Florida, Tennessee, and Rhode Island were, according to policy implementation timelines, subject to high-stakes professional sanctions as a consequence of unsatisfactory evaluations. Given the gravity of these summative outcomes, it is concerning that a substantial proportion of music educators did not have a clear understanding of teacher evaluation policy with regard to summative objectives and personal career implications. The Washington Educators Association, as referenced by Duke (1995), described unclear evaluative outcomes and processes as creating a "risky, threatening environment for school employees" (p.

145). Such perceived threats may influence music educators' receptiveness to new teacher evaluation policies, potentially contributing to skepticism of evaluation's overall legitimacy.

*Evaluation purpose: What is versus what should be.* In addition to reporting the objectives associated with their evaluation process, music teachers expressed their evaluative function preference by indicating the extent to which certain objectives *should be* emphasized. Participants' perceptions of existing teacher evaluation functions represented *what is*, while preferred functions represented *what should be* with regard to the purpose of the evaluation process. Corroborating previous research (Flores, 2012; Gratton, 2004; Kelly et al., 2007; Nolin et al., 1994), participants believed teacher evaluation *should be* more formative and developmental in nature (i.e., to guide improvement of teaching, recognize and reinforce excellent teaching, and help teachers focus on student learning) than summative or punitive (i.e., to inform high-stakes personnel decisions). This finding again implies that music educators may be more drawn to intrinsically oriented evaluation systems emphasizing autonomy and professionalism.

Participant reports of *what is* and *what should be* were compared for each evaluation objective to gauge the fit between policy and music educator preference. With the exception of using teacher evaluation results to determine teachers' merit pay, participants' responses suggested they believe all proposed evaluation objectives *should be* considered to a greater extent – which may not be practical from the standpoint of time and effort constraints that are part of any large-scale evaluative program. However, in considering the relative emphasis (rank ordering of mean evaluations) ascribed to each of these evaluation objectives, music teachers' lived and idealized notions of evaluation objectives were largely aligned. It may be that music teacher responses regarding *what should be* reflect a more general belief that evaluations are not

adequately or comprehensively capturing teacher performance.

Music educators' apparent disinterest in merit pay as an evaluation objective may be a function of the limited scope of pay-for-performance policies in education, or even the imbalance of merit-based opportunities afforded to teachers in non-tested subjects such as music (Elpus, 2011). Moreover, RttT policy implementation may draw music educators' immediate concerns toward more pressing components of the evaluation process. For example, music educators may be more focused upon tracking student growth instead of the ensuing rewards or sanctions that could potentially impact salary, job security, or career longevity. Alternatively, participants may simply lack support for merit pay as a mechanism of teacher compensation, regardless of how teacher evaluations factor into merit pay decisions.

***Measures of teacher performance.*** Participants reported that classroom observations, both formal and informal, were considered to a greater extent than other indicators of teaching effectiveness, including student performances, student achievement data, student survey data, and records of professional development activity. The emphasis on classroom observations may reflect participants' familiarity and comfort with the observation component, as it has traditionally been the sole indicator of teacher effectiveness. However, relying primarily upon classroom observation data may be problematic for music educators, particularly if evaluators conducting classroom observations have little to no training in music, or when classroom observations significantly outweigh evaluators' consideration of other data sources as part of a more robust, multiple measures approach to assessing teaching effectiveness.

A substantial proportion of participants whose overall evaluation incorporated student achievement data reported that *non-music* standardized tests scores were considered part of the student growth component. To compound the issue, in many cases, participants reported their

evaluations were based upon non-music test scores for students not even enrolled in music classes. Most testing experts would contend that scores reflecting student achievement in subjects like reading, writing, and math are not a valid representation of a *music* teacher's effectiveness, because music teachers have limited control and direct impact where learning in those subject areas is concerned (Buckley & Marion, 2011). RttT states in this study have adopted various approaches to measuring student growth in music, including formal assessments of knowledge and performance, teacher portfolios, and Student Learning Objectives (SLOs). While Rhode Island currently implements SLOs, the three remaining states in this study were, at the time of data collection, either implementing or developing alternative assessments to gauge teacher contributions to student growth *in music*. Despite these recent initiatives, participants' responses still suggest that non-music student growth measures are factored in to music teacher evaluations.

Interestingly, nearly one-third of participants indicated that student achievement and growth were *not at all considered* in their most recent evaluation. Still others were not sure what type of role student achievement data played in their evaluations. According to RttT's "Great Teachers and Leaders" criteria, student growth should account for a *significant portion* of teachers' overall evaluation (United States Department of Education, 2011). Though individual states' implementation timelines were staggered, Florida and Tennessee instituted student growth considerations during the 2011–2012 school year, with Rhode Island following in 2012–2013. Only North Carolina delayed consideration of the student growth component until the 2015–2016 school year, collecting longitudinal data on student growth during the three school years leading up to 2015–2016. North Carolina music teachers accounted for less than one-half of participants who believed student growth and achievement was not part of their overall

evaluation, signifying that a number of music educators teaching in states purportedly considering the student growth component in teacher evaluations were unclear as to whether they would be evaluated on the basis of student achievement.

*Measures of teacher performance: What is versus what should be.* In terms of various measures of teacher performance, participant reports of *what is* and *what should be* considered in the evaluation process were compared. Music teachers indicated that formal and informal classroom observations *should be* given greatest consideration, which suggests a certain level of comfort with the status quo approach to measuring teacher effectiveness on the basis of observational data and administrator evaluations. Participants in this study also expressed the opinion that portfolios *should be* given more emphasis (i.e., should be considered to a moderate extent while presently considered only slightly or not at all) in the teacher evaluation process. The favorable response to portfolio implementation alongside other measures of teacher performance corroborates findings by Attinello et al. (2006), who determined that portfolios contributed to a more complete perspective on teacher effectiveness when compared with observations alone. Support for portfolios may reflect a belief that teachers want their evaluations to be more authentic – reflecting what they actually do with students in their classroom as professional educators (Robelen, 2013). Participants were less inclined to endorse the consideration of student performances (i.e., concerts or contest ratings), student achievement data, and student survey data in the evaluation process. Perhaps this reflects a belief that such measures are, at best, proxies for teacher effectiveness given the number of factors that may influence student performance and opinions. Ultimately, participants expressed support for various measures of teacher performance, including but not limited to classroom observations, which aligns with the multiple-measures approach embedded within the “Great Teachers and

Leaders” guidelines for RttT funding. As Barrett (2011) stated, “Quality in music teaching is a complex and sophisticated process not easily captured by platitudes and checklists” (p.1). It is therefore understandable that participants were interested in developing a multi-dimensional, comprehensive understanding of their professional practice through multiple measures of teacher performance.

***Evaluator role.*** The vast majority of music educators in this study identified their primary evaluator as a building administrator with no music expertise. In fact, less than ten percent of participants reported having a primary evaluator with music expertise. In open-ended items, evaluators’ lack of music expertise was a recurring theme, with music educators expressing concern about their evaluators’ ability to effectively assess music classroom teaching. As one participant contributed, “Evaluators who have no experience in a performance ensemble don’t understand what’s happening when they observe a class, and they can’t effectively critique what they’ve witnessed.” Building administrators lacking music expertise can provide generic feedback regarding instructional delivery or classroom management techniques, and while this feedback may be useful, it may not adequately address a music teacher’s need for critique specific to content area pedagogy. Shulman (1987) referred to teachers’ unique professional understandings as *Pedagogical Content Knowledge* (PCK), a “special amalgam” that:

...represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted into the diverse interests and abilities of learners, and presented for instruction. (p. 8)

If evaluators are unable to provide specific feedback that helps shape and refine music educators' PCK, the evaluative process may be considered less valuable with regard to professional growth and development.

Very few participants reported that district music administrators or other music teachers were involved in their evaluation, let alone that these individuals were primarily responsible for determining effectiveness ratings. This may be due to the limited number of district-level music supervisor positions in existence today, as well as the multiple responsibilities assigned to music supervisors. There may also be a lack of structural and financial support for peer mentoring and evaluation among teachers more generally. As music educators favored formative evaluation objectives emphasizing professional development, restricting involvement and responsibility for music teacher evaluations to building administrators limits opportunities for professional growth specific to the content area. These missed opportunities might cause music educators to view evaluations as impersonalized, lacking purpose, and demotivating. Music educators might also develop mistrust for their evaluators, born from concern that evaluative outcomes may too often reflect misinterpretations of classroom events. These consequences, of course, have short-term implications, such as dissatisfaction with the evaluation process. However, over time, missed opportunities for comprehensive, multi-perspective evaluations yielding meaningful and content-specific feedback can contribute to lowered professional morale, potentially impacting overall commitment to the profession.

**Section summary.** Goe, Holdeheide, and Miller (2011) assert that comprehensive teacher evaluations systems must have “explicit, well-defined goals” (p. 9). In this study, certain critical elements of state-level teacher evaluation policy may not have been made explicit or transparent to study participants. Perhaps building and district administrators charged with teacher

evaluation receive inadequate information and training to educate teachers on the evaluative process. There may also be poor communication between administrators and the teachers themselves. Another potential explanation for participants' lack of clarity could be the time it takes for newly enacted educational policies to reach and affect teachers in non-tested subjects (Madaus, Russell, & Higgins, 2009). Because RttT teacher evaluations were designed with little consideration for teachers in non-tested subjects, it may take even more time for educators in areas such as music and art to fully experience new policies. Without understanding evaluation objectives and expectations, it is logical that music educators may be frustrated with or confused by the evaluation system as a whole. This frustration may be amplified when administrators without music expertise facilitate evaluations, or when non-music standardized test score data factors into music teachers' overall evaluations.

#### **Perceptions of the music teacher evaluation process.**

*Evaluation qualities.* With regard to practical qualities of the evaluative process, music educators in this study were, on average, slightly dissatisfied with the teacher evaluation process. Participants were divided with regard to evaluation clarity, with a slight majority (54%) believing the process was clear, while the rest of music educators were either unsure or unclear on evaluative expectations. Kelly et al. (2007) determined that perceived clarity of the evaluative process was a major predictor of teacher satisfaction with overall evaluations. Because Kelley and colleagues found perceived clarity to also be a strong predictor of job satisfaction and professional motivation, stakeholders might focus attention and resources upon more effectively disseminating evaluation policy expectations to *all* educators. As an additional benefit, clarifying evaluative expectations may also positively impact teachers' professional development and growth (Delvaux et al., 2013).

Some participants expressed the opinion that the current evaluation system *should not* be continued, even though implementation had taken place for three years or less depending on the state in which a teacher worked. While a few music educators in this study craved a new or different evaluation process, participants described few alternatives. Overall, the willingness to abandon RttT-driven teacher evaluation policies suggests an ultimate lack of buy-in on the part of music educators, which may determine whether or not associated reforms will have staying power in the long term (Schmidt & Datnow, 2005; Spillane et al., 2002).

Participants reported that teacher evaluation did not provide an accurate depiction of their teaching performance, nor was their performance equitably or fairly assessed. These reports may be rooted in concerns about the evaluation process as discussed earlier in this chapter (e.g., validity of non-music student growth data, evaluator area of expertise). According to Kelly and colleagues (2007), fairness is also a strong predictor of professional satisfaction. Open-ended responses highlighted music educators' concerns about fairness, with one participant writing, "[The teacher evaluation process] does not completely assess or measure all the things that music teachers do in their classrooms." To achieve the collective buy-in necessary to sustain new evaluation policies, it is critical that teachers believe those policies are fair and unbiased. Targeting participants' major areas of concern, such as evaluator area of expertise, may be the first step in addressing issues of fairness and equity. Developing teachers' collective buy-in to RttT evaluation policy will prove particularly critical to sustaining associated initiatives when allocated federal grants run out, and states must determine whether they will continue funding more robust teacher evaluation systems.

While some evidence suggests the amount of teaching experience may influence teachers' reaction to teacher evaluation processes (Kimball, 2002), no such relationship was

found in this study. In other words, both early-career and veteran teachers demonstrated similar attitudes toward the evaluation process. This finding may suggest that, to a degree, RttT policies objectively assess how teachers – across all levels of experience – are evaluated, thereby muting experience-related differences in perceptions. This result may also imply that early-career teachers do not envision becoming career-long supporters of RttT evaluation policies, further threatening the long-term viability of RttT reform.

*Evaluator characteristics.* In responding to questionnaire items that addressed the primary evaluator’s characteristics and competence, music educators were generally of the opinion that their evaluator had considerable teaching experience, did a thorough job while being relatively fair and objective, and established good interpersonal relationships.

Music educators’ beliefs about their primary evaluators as expressed through responses to scaled questionnaire items were often illuminated by the information-rich comments participants provided about their evaluators in response to open-ended items. As one participant contributed, “[My evaluator] is open-minded when we discuss how general classroom strategies are adapted in the music classroom.”

Open-ended responses also provided perspectives that contrasted closed-ended item results. One participant noted that, with multiple administrators involved in the process, evaluations were rarely consistent. Another participant asserted the evaluator allowed “personal differences to get in the way of his objective evaluation.” These concerns, while valid, represent more global apprehensions that may also be embraced by general classroom teachers.

Participants were less inclined to trust their evaluator to accurately assess teaching performance. Music educators were also less likely to view evaluators as qualified to assess their teaching. Many narrative comments elaborated the general concern that primary evaluators

lacked music expertise. Researchers have demonstrated that teachers have higher regard for evaluators with deep content knowledge, as they can contribute more practical advice for professional improvement (Delvaux et al., 2013; Goe et al., 2012; Stiggins & Duke, 1988; Wise, et al., 1984; Zimmerman & Deckert-Pelton, 2003). In this study, music educators' evaluator-related concerns focused largely on the classroom observation component of the teacher evaluation process, with many music educators commenting on evaluators' lack of familiarity, understanding, or expertise in matters of typical music classroom behaviors and practice. As one participant commented, "I sometimes feel administrators don't know what to say when it comes to music evaluations...[my administrator] has no idea how a music classroom is run."

Because participants also indicated that classroom observation data were considered to the greatest extent in their overall evaluations, concerns regarding evaluators' music expertise, as well as the accuracy and validity of the effectiveness ratings they subsequently produced, pose important implications for teacher morale and the sustainability of the RtT teacher evaluation initiative. Lowered morale and commitment to the profession were evident in participant comments, including, "All of the seasoned veteran teachers I know cannot wait to get out [of the profession]," and, "I no longer tell my high school students to go into education."

Despite music teacher concerns about their evaluators, a small percentage of participants were appreciative of their evaluators' attempts to compensate for lack of music-specific expertise. Open-ended responses indicated that music educators valued evaluators' questions about their unique classroom expectations, as well as when evaluators posed questions regarding music-specific content. In other words, music teachers appreciated an evaluative approach that fostered professional dialogue and mutual understanding, rather than one that positioned music teachers and administrator-evaluators as assuming antagonistic roles. This finding is encouraging

and substantiates the work involved in developing the *Workbook for Building and Evaluating Effective Education in Music*, a tool created by the National Association for Music Education (NAfME) in an effort to make music classroom observations conducted by administrators more viable and useful. Previous research confirms that teachers crave more interactions and dialogue with their evaluators (Campbell, 2013; Zimmerman & Deckert-Pelton, 2003). Music expertise aside, continued research on teacher and evaluator relationships may shed light on strategies by which building administrators and music educators can work together to enhance the utility of annual evaluations.

**Section summary.** In this study, participants were generally dissatisfied with the teacher evaluation process. Music educators' concerns regarding the accuracy and equitability of RttT-driven evaluation practices centered upon apprehensions about the primary evaluator's lack of music expertise. These issues, combined with music educators' general desire to abandon RttT evaluation practices, complicate RttT's long-term reach in K–12 schools. When teachers do not perceive favorable returns on their investments, it impacts their commitment to a given initiative (Gratton, 2004). This reality holds true for music teachers surveyed in the current study, as evident in their general dissatisfaction with RttT teacher evaluation policies. In keeping with Gardner's (2010) model of music teacher retention, turnover and attrition, such negative perceptions of the workplace may impact overall job satisfaction, and in turn, whether or not music educators intend to stay in the profession. As evident in open-ended responses, RttT policies may motivate some music educators to actively deter others from entering the profession, as well. Ultimately, negative reactions to the teacher evaluation process may have long-term implications on the amount of quality music educators that choose to pursue and remain in the profession.

**Perceived benefits and impact of music teacher evaluation.** Participants indicated the degree to which they believed the evaluation process benefitted various aspects of their teaching, including teacher effectiveness and student achievement. Of all potential professional benefits, participants believed the evaluation process most enhanced their professional self-reflection. As one participant contributed, “[The evaluation process] really encouraged me to rethink the way I taught.” Although self-reflection is a critical component of an effective evaluation cycle, self-reflection alone does not necessarily directly improve teacher performance or student achievement. Additional research might highlight practical applications of self-reflection associated with RttT-oriented evaluation processes, to better understand if and how self-reflection translates to improved practice in the music classroom. Furthermore, though self-reflection was reported as the strongest benefit to teacher evaluation when compared with other potential benefits, on average, participants still believed the evaluative process had a slightly negative impact upon professional self-reflection. Participants frequently described the time demands associated with evaluation expectations, suggesting music educators may have felt too stretched to engage in regular, meaningful reflective practice. In the context of this study, perceived benefits to self-reflection were ultimately overshadowed by music educators’ concerns regarding teacher evaluation’s impact upon other key issues.

Music educators expressed notable reservations, ambivalence, and even doubt as to how they and their students benefitted from RttT’s revamp of the teacher evaluation system. These results corroborate research by Schumacher (2010), who determined that teachers did not always perceive the evaluative process as having a positive influence upon professional and educational outcomes. With regard to the perceived lack of benefit upon *teaching practice*, results suggested that, in spite of policy intentions, RttT-driven evaluations might not have afforded teachers

appropriate opportunities for professional growth. Limited opportunities for teachers' growth and development may be rooted in evaluators' lack of content expertise or an overall shortage of time and resources. Still, music educators perceived the evaluation process as primarily formative in nature, revealing a disparity that may further illuminate why music educators expressed discontent with the evaluation process.

In terms of benefits to *student growth*, participants were least agreeable with the notion that evaluation policy contributed to improved student learning, which, according to Hallinger et al. (2014) represents a primary goal of the teacher evaluation process. The accountability movement is driven by the goal of ensuring student achievement; as such, it is worrisome that music educators do not see teacher evaluation – conceived to enhance teacher effectiveness defined in part by student achievement – as accomplishing that goal. Perhaps this incongruity is attributable to the ways in which *music* student achievement is “measured” in RttT evaluation systems. Music educators may perceive few connections between the evaluative process and student learning when student growth data fails to reflect music-specific progress.

Participants also indicated the degree to which they believed the evaluation process impacted various daily classroom practices, such as the type of instructional materials used or the way in which instructional time was utilized. In responses to closed-ended items, participants reported the evaluation process had a slightly negative effect on most aspects of their daily routine. Music educators demonstrated concerns regarding evaluation's impact upon their autonomy and decision-making with regard to appropriate assessment practices and making independent choices. Such threats to professional autonomy may be the essence of why music teachers demonstrate slightly negative feelings, overall, toward new evaluation policies. Day, Elliot, and Kington (2005) describe the implications of policy change upon professional

autonomy and identity:

Policy changes and reformist imperatives have left many teachers themselves feeling confused about their professional identity, the extent to which they are now able to use their discretionary judgment—arguably at the heart of their sense of professionalism—and about their capacity to carry out the responsibilities associated with their new performativity identities, which challenge traditional notions of professionalism and professional purposes and practices. (p. 566)

Despite the general negative response to evaluation's influence upon classroom practices, open-ended responses demonstrated more favorable reactions with regard to how teacher evaluation processes influenced classroom practice. As one participant stated, "The evaluation process pushes me to provide high quality instruction, meeting students at various levels." While some participants sought to maintain autonomy and control in their classroom, other music educators seemed to appreciate the structure and standards embedded in the evaluation process.

Open-ended responses highlighted additional details relevant to teacher evaluation's impact upon music educators and music student learning. More participants believed that RttT-oriented teacher evaluation practices had *no* apparent impact upon their teaching practices or student learning, when compared with participants who believed teacher evaluation had a negative impact on similar factors. While these findings support some of the neutral responses evident in quantitative results, the perceived lack of impact may also be related to the reality that, despite implementation timelines, many music educators have yet to live evaluation reforms as they were intended.

An additional theme related to classroom impact was participants' sentiment that the teacher evaluation process was a "hoop through which they had to jump". Concerned participants suggested the evaluation process inadvertently drove them toward inauthentic practice in order to meet evaluative expectations, only to revert back to typical instructional behaviors when not under the watchful eye of the evaluator. The bureaucratic demands associated with current teacher evaluation processes may, indeed, drive music educators to feel like "an object of bureaucratic scrutiny" (Wise & Darling Hammond, 1985, p. 28), thereby pushing them to make unconventional adjustments to instruction to meet expectations.

Reault (1998) suggests that the context of teacher evaluation (e.g., bureaucratic versus autonomy-oriented) may color teachers' impressions of its overall value. If music educators perceive the teacher evaluation process to be a chore rather than a vehicle for professional growth, then the process may be both poorly designed and poorly executed. The "hoop-jumping" mentality essentially ensures a given policy will not gain long-term traction with the music education profession.

***Benefits of teacher evaluation: Group differences.*** In examining group differences across music educators' perceptions of various aspects of the teacher evaluation process, a few significant differences surfaced. For example, elementary music teachers saw the evaluation process as more beneficial than high school music teachers. Similarly, general music teachers believed the evaluation process was more beneficial to them than did instrumental music teachers. These two findings seem logical, as the majority of elementary music educators teach general music, and instrumental music is more prevalent at the secondary level. An additional group difference existed when considering gender – females were more likely to find evaluations to be beneficial than males. Connecting back to the level (e.g., elementary, secondary) and area

(e.g., general music, instrumental) findings, this result is complementary, as the majority of elementary general music teachers are, in fact, female. However, these results should be interpreted with caution, as they reflect potential sampling bias complications, insofar as elementary music teachers were not proportionally represented in the response sample.

Arguably, the elementary general music classroom might be the scenario that most closely imitates the traditional, non-music classroom. Typically, the general music classroom is non-selective, including all students in an intact grade-level class, and in many cases, general music is a requirement for elementary students. Conversely, many secondary music classes are offered as electives, and in some circumstances, may require financial capital for students to participate (e.g., instrument purchase). Therefore, secondary music classrooms may address a more select student population. Class sizes in elementary general music classrooms are also more reflective of a traditional, non-music classroom, while enrollment in secondary instrumental music performance ensembles, for example, could exceed 50 students. Furthermore, in working with younger children from more diverse populations, elementary general music teachers may also develop a unique ethic of care that predisposes them to acceptance and support, including within the context of new evaluation efforts. The distinct differences between secondary instrumental music classrooms and traditional classrooms may explain the comparative discontent those participants expressed regarding RttT-driven teacher evaluation policies. These group differences might highlight a practical starting point from which adaptations to the evaluative process might first be considered.

**Section summary.** Overall, music educators surveyed in this study believed RttT-driven teacher evaluation practices had few benefits. Participants also reported that evaluation processes had a slightly negative impact on many aspects of independent classroom decision-making. This

sentiment may reflect a belief that the current evaluation expectations threaten or impinge upon music teachers' values and autonomy (Muncey & McQuillan, 1996). Teachers who perceive their work as "uncertain and complex" (Rosenholtz, 1987, p. 537) naturally desire more autonomy. Indeed, prior research highlights the unique nature of the music classroom (e.g., Baker, 1981; Shaw, 2013; Taebel, 1990b), while several other researchers have noted the importance of teachers' professional autonomy, particularly in the context of reform (Del Savio, 1992, Perie & Baker, 1981; Reault, 1998). It is possible music teachers are unsure about how to effectively fit the evolving teacher evaluation system to their practice in a meaningful manner. A desire for autonomy may be born from disparate conceptions of effective music teaching held by music educators, administrators, and policymakers (Rosenholtz, 1987).

Strict evaluation policy regulations may imply that stakeholders mistrust teachers' professional judgment, further fueling music educators' negative reaction to current initiatives. This mistrust is amplified when high-stakes sanctions are associated with teacher evaluations, as the "unquestioning compliance with agency directives" may drive music teachers to compromise their instructional integrity – as demonstrated, for example, by "putting on a show" or "jumping through hoops" (Darling-Hammond, 1990b, p. 31). Darling-Hammond suggests that deregulating requirements of teachers' daily classroom practice, while simultaneously upholding higher standards through which teachers enter the profession, may alleviate patterns of mistrust, thereby contributing to improved teacher effectiveness.

**Music education as a profession.** Participants largely believed that the teacher evaluation process diminished the status of music education as a profession, with one person describing feelings of being "an afterthought" in the process. More specifically, some music teachers reported a deflated sense of professional morale due to RttT-oriented evaluation

demands, feeling their content area was undervalued in comparison with tested subjects. Several researchers have found similar connections between reform implementation and teachers' emotions, commitment, and vulnerability (Lasky, 2005; Kelchtermans, 2005; Schmidt & Datnow, 2005).

While roughly one-third of participants felt that the evaluation process diminished the status of music education, a smaller percentage of participants believed that uniform teacher evaluation practices helped elevate the status of music education from a professional standpoint, as it made music "equal" to other subjects. By being held to the same standard as teachers in other content areas, many participants were encouraged by the new evaluation expectations. Crowe (2008) points out that professional communities are born from "shared norms, training, working practices, and regulatory mechanisms" (p. 990). The "shared norms" of evaluation expectations across different content areas effectively communicate that teaching is teaching, no matter the discipline. Though imposing uniform standards across diverse teaching disciplines may be problematic, the practice may also have some merit in "leveling the playing field" for music teachers, enhancing their professional legitimacy in the eyes of teachers from other content areas. In fact, beyond their concerns for the music education profession specifically, some participants expressed worry regarding teacher evaluation's impact upon the education profession as a whole. In other words, some participants recognized the broader implications of evaluation reforms, outside of the music classroom. These more global concerns may highlight an opportunity for teachers across disciplines to collaborate on more appropriate approaches to teacher evaluation that enhance "shared norms" in a manner that better upholds the education profession at large.

In responding to the prompt regarding teacher evaluation's impact upon the status of music education as a profession, many participants alluded to or revisited concerns similar to those expressed in the first open-ended item on the impact of evaluation upon teaching practice and student learning. For example, participants reiterated concerns about the time demands associated with the evaluation process, as well as concerns with the evaluator's expertise and competence. These overlaps may be an effect of participants' train-of-thought continuing on from the first open-ended item. However, unlike the first open-ended item, a few participants expressed concerns regarding their *own* professional status, particularly in terms of attaining tenure or being dismissed from their position.

Considering the high-stakes sanctions associated with RttT-oriented teacher evaluation practices, it is understandable that some music educators might operate from a self-preservation standpoint with regard to professional status. This self-preservation approach reflects a mentality of being on one's own in the evaluation process – a sentiment corroborated in participants' comments about “relying on one's self” to make the most of a difficult situation. Researchers have demonstrated that music teachers often experience elevated feelings of isolation in the profession (Sindberg, 2011; Sindberg & Lipscomb, 2005). These feelings may be amplified under the stressful conditions of a new evaluation system that may not appropriately fit their professional development needs or threaten professional autonomy in the classroom.

## **Implications and Recommendations**

### **Suggestions for policymakers.**

*Teacher evaluation design emphasis.* Teacher evaluation policymakers are faced with the challenge of designing an appraisal system that is uniform enough to facilitate consistent, objective results, while also being flexible enough to accommodate educational practices in non-

traditional or non-tested subject area classrooms. With the majority of the teaching population specializing in non-tested subject areas, it appears as though the design emphasis is misplaced upon the classroom teachers in the minority (i.e., math and reading content specialists). If the United States continues with its push toward standardized measures of teacher evaluation, more consciously considering the needs of the *majority* of teachers may help policymakers develop teacher evaluation systems that are more applicable and useful all around. This approach would necessitate a paradigm shift, refocusing public attention away from math and reading as the most “important” subjects, toward a broader concept emphasizing comprehensive and diverse education that values other subject areas as equals.

***Localizing control.*** Another issue with RttT teacher evaluation is the indirect federal control that is being exercised in an arena that *should be* chiefly within state and local jurisdiction. True, RttT grants were optional, and states had the opportunity to develop and tweak their evaluation systems to suit their unique needs. Ultimately, however, in order to be eligible for RttT federal funding, state reform plans had to adhere to certain federal guidelines. This approach effectively holds local education agencies as financial hostages to the federal government, as funding is predicated upon compliance with prescribed policies. Because teaching is a nuanced, contextualized endeavor, standardizing its evaluation at the national level may not provide the most authentic teacher appraisals. Refocusing evaluation systems at the building, district, or regional level will allow for a more refined approach to teacher evaluation that better serves the professional development needs of teachers, enhancing students’ experience in the classroom.

**Suggestions for music education leadership.**

***Mutual accountability.*** Leaders in music education must hold policymakers accountable

for the repercussions of teacher evaluation policies that adversely impact the profession. Aguilar and Kapalka Richerme (2014) note that arts educators may perceive “problems resulting from teacher evaluation policies as inadvertent” (p. 118), when, in fact, policymakers have the responsibility to be aware of how policies might play out for a variety of players in diverse contexts. The authors go on to suggest:

Rather than passively accepting the system that is presented, teachers who view policy actors as intentional agents may feel empowered to become involved with the policy process, hopefully speaking up and offering suggestions about how to make the system more fair and transparent for all teachers, especially those in nontested subjects. (p. 118)

As such, if policymakers are going to hold music teachers accountable to certain evaluative expectations, so should music education leadership hold policymakers accountable for developing and implementing teacher evaluation policies that benefit music teaching and learning.

***Engaging stakeholders.*** NAFME has made several focused efforts to address the growing issues surrounding music teacher evaluation. In addition to their position statement describing recommendations for music teacher evaluation (National Association for Music Education, 2011<sup>a</sup>), NAFME released an advocacy guide designed to facilitate music educators’ understandings of teacher evaluation policies and the ensuing implications for the music classroom (National Association for Music Education, 2011<sup>b</sup>). The *Workbook for Building and Evaluating Effective Education in Music*, released by NAFME in 2013, represents an additional effort to effect positive change in music teacher evaluation practices. The Society for Music

Teacher Education (SMTE), a subsidiary of NAFME, has also compiled resources specific to music teacher evaluation, which provide K–12 music teachers and other stakeholders information regarding state-level teacher evaluation policy implementation.

However, results from this study suggest that recommendations and efforts made at the national level may be lost in practice at the state and local level. Music education leadership has a responsibility to ensure these strategies are effectively communicated to K–12 teachers and applied on-the-ground. As such, in keeping with Aguilar and Kapalka Richerme’s (2014) call to action, music education leadership must strengthen its working relationship with classroom music teachers. To facilitate these connections, music teacher educators might effectively bridge the gap between national initiatives and local implementation. However, it is important to ensure this influence extends beyond the immediate geographic reach of music teacher preparation programs.

*Measuring music-specific student growth.* The next challenge is to determine how to best measure music student growth as part of the music teacher evaluation process. Based on current policy and practice found in RttT states, portfolio-based systems, state- or district-level written tests of music knowledge, and state- or district-level performance assessments seemed to be considered increasingly viable as discipline-specific growth measures. Long-term, and for support to develop outside of the music educator ranks, the reliability and validity of these approaches must be established through careful development, implementation, and analysis. In doing so, teacher evaluations may better account for music teachers’ direct influence on their students’ learning, in contrast to “the host of environmental factors or student attributes” (p. 166), which Ravitch (2010) asserts contributes to student achievement school-wide.

Various state-level initiatives, such as Florida’s Performing Fine Arts Assessment Project

and Tennessee's Fine Arts Growth Measure, already reflect meaningful action taken by music education leadership to ensure more accurate assessment of music student growth, thereby enhancing the validity of music teacher evaluations. However, *standardized* assessment of student growth in music poses practical challenges, while also raising important questions about how effective music teaching is defined. Perrine (2013) queries whether or not standardized assessments are even appropriate for the performing arts, describing specific challenges for music:

A standardized test might indicate that a student understands music conceptually, but can say nothing about whether a student can actually *make* music. By extension, such a test cannot reveal whether teachers are being successful in educating their students in how to produce high-quality music together. (p. 43)

Perrine's comments highlight the distinctiveness of music as a subject area (i.e., knowledge *and* skill acquisition, individual *and* collective performance) such that depending on standardized music assessment data alone to provide a complete and reliable measure of teacher performance would be problematic. The process of using test scores to represent student growth and inform teacher evaluations may also diverge from some music teachers' professional pedagogies and personal philosophies. Music teachers tend not to gauge their own effectiveness by objective scores, instead referring to other indicators of student progress. Music education leadership must explore mechanisms for determining student growth in music that bridge the gap that currently exists between music teachers' personal philosophies and the real need to define and measure music student achievement.

*Observations by music experts.* Those best suited to evaluate music teachers when collecting classroom observation data, in a manner that promotes professional growth and development, are other music teachers. As such, local, regional and state level music education leadership should develop sustainable and transferable models for music educators to engage in peer evaluation. This could entail funding short-term substitute teachers so that expert music educators could leave the classroom and travel to evaluate teachers in other buildings. Developing a system of peer review could also enhance music teachers' sense of professionalism. Furthermore, such observations might serve as opportunities to better inform non-musically trained building administrators. For example, a building principal could observe a music classroom alongside a trained music educator, who helps point out successes and challenges as they occur, thereby "training" that administrator on what to notice, when, and why when observing a music teacher.

Darling-Hammond (2014) and Wise et al. (1984) advocate for including master teachers with subject-specific expertise in the evaluation process, particularly for new teachers and teachers needing extra support. Study participants reported classroom observations are presently the most heavily weighted component in music teacher evaluations, and music educators expressed a clear desire that classroom observations continue to be a prominent part of a more comprehensive set of teacher effectiveness measures. Therefore, ensuring that a content-area expert conducts at least *some* observations would likely yield more reliable and valid reports of teaching effectiveness, while also providing music teachers with more meaningful feedback (particularly when that feedback is provided in a timely fashion). This process can, in turn, more effectively shape the professional growth of music educators. There are costs (personal, time, financial) associated with implementing a peer-based or blended (peer and administrator)

approach to conducting classroom observations within a rigorous teacher evaluation framework. Yet, sharing classroom observation evaluation duties with master teachers can also alleviate time demands that burden administrators, freeing them to consider other measures of teacher performance with greater discernment.

**Suggestions for classroom music educators.** Music teachers have a responsibility to educate themselves on policies that impact their classroom. In this study, a notable number of music educators reported confusion regarding evaluation expectations. While communication breakdowns in local leadership may contribute to this lack of clarity, there exist numerous resources (such as those suggested by NAFME and SMTE, described earlier) that music educators can access and apply to their practice. Rather than simply endure these new policies, music teachers must be active, educated agents in the evaluative process, advocating for clear and appropriate evaluations.

Beyond self-advocating, music educators should keep careful, detailed notes on the evaluation process, including the manner in which it is conducted, the evaluators responsible for determining music teacher effectiveness, and the type and source of student growth data considered in overall evaluations. Examining the evaluative process in detail may highlight key areas for future improvement, and careful records may help safeguard music teachers subject to sanctions based upon evaluative data beyond their control. Detailed records of evaluative practice can also help stakeholders gauge the consistency with which policies are enforced among music teachers at the building and district level.

Finally, in the face of accountability expectations, classroom music teachers must reconsider how success is defined in music education. While stereotypical notions of winning trophies at contests have historically characterized what some music teachers consider to be

“success,” the broadening curricula and increasingly diverse student population demand a more creative and progressive take on what makes a successful teacher, student, and music program.

### **Study Limitations and Recommendations for Future Research**

To better contextualize significant findings in this study, it is important to recognize the various limitations of this study as related to sampling, measurement, methodology, and data analysis. These limitations also illuminate potential areas for future research:

#### *Sampling Concerns*

- Because NAFME does not require certain demographic information from its members, such as area of specialty, it was difficult to ascertain whether or not a representative sample of participants contributed to the study. To help facilitate more comprehensive research in the future, NAFME might consider requiring such data from its members.
- Limiting the population sample to MEA members may have impacted the generalizability of results. In future studies, researchers might draw from a more representative sample of music educators by working directly with state departments of education.

#### *Methodological Concerns*

- Due to varying terminology associated with new teacher evaluation processes across multiple states, it was difficult to determine appropriate language for certain items, in order to ensure participants in each state unequivocally understood the intent of each item. Future researchers should take varying terminology into strong consideration when designing survey instruments intended to gather information on state policy.

- To limit the burden upon MEA representatives who agreed to serve as the middlemen for relaying study information to state MEA members, a single follow-up email was sent to study participants. An additional follow-up email may have improved the overall response rate. In future studies involving online questionnaires, researchers should strive to send at least two follow-up emails.
- The two extended open-ended items occurred in succession in the online questionnaire, which potentially influenced the recurring reference to similar themes (e.g., concerns about the evaluator). Separating open-ended items in the questionnaire design may help to control this overlap.

#### *Analytical Concerns*

- The varied response rates from each of the four participating states limited the possibilities of facilitating cross-state comparisons. Future researchers who achieve more representative response rates from multiple states might explore group comparisons in greater depth, so as to better identify what works in terms of music teacher evaluation, for whom, and why.

#### *Additional Recommendations*

- Future researchers examining teacher evaluation practices should strive to uncover systems that music educators find professionally satisfying, so as to determine ways to replicate such systems in other contexts.
- By qualitatively exploring the experiences of career music teachers who have transitioned through multiple evaluation processes, researchers may gain a more in-depth perspective of

how evaluation policy plays out for music educators, as well as how *long* such policies take to manifest themselves in the music classroom.

## **Conclusion**

This study revealed that music educators have mixed feelings about the professional utility and fairness of teacher evaluation practices in RttT states. Ultimately, in order for music educators to draw professional meaning from teacher evaluations, and to subsequently enhance music student learning, there must be a shift to ensure music teachers find the evaluative process adequately clear, fair, practical, sustainable, and useful. Policymakers may find evaluation systems that fail to meet these core professional needs struggle to exercise meaningful, lasting impact upon teacher quality.

## References

- Abeles, H.F. (1975). Student perceptions of characteristics of effective applied music instructors. *Journal of Research in Music Education, 23*(2), 147-154.
- Aguilar, C.E., & Kapalka Richerme, L. (2014). What is everyone saying about teacher evaluation? Framing the intended and unintended consequences of Race to the Top. *Arts Education Policy Review, 115*(4), 110-120.
- Attinello, J.R., Lare, D., & Waters, F. (2006). The value of teacher portfolios for evaluation and professional growth. *National Association of Secondary School Principals Bulletin, 90*(2), 132-152.
- Baker, B.D., Oluwole, J.O., & Green, P.C. (2013). The legal consequences of mandating high takes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education Policy Analysis Archives, 21*(5), 1-67.
- Baker, P.J. (1981). *The development of music teacher checklists for use by administrators, music supervisors, and teachers in evaluating music teaching effectiveness*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses. (UMI No. 8201803)
- Ballou, J. (2011). Open-ended questions. In P.J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (547-549). Thousand Oaks, CA: Sage.
- Barrett, J.R. (2011). Judging quality and fostering excellence in music teaching. *Journal of Music Teacher Education, 21*(1), 1-6.
- Bean, A.G., & Roskowski, M.J. (1995). The long and short of it. *Marketing Research, 7*(1), 20-26.
- Bentley, R., & Rempel, A. (1970). *Manual for the Purdue teacher opinionaire*. West Lafayette, IN: University Book Store.

- Berliner, D.C., & Glass, G.V. (2014). *50 myths and lies that threaten America's public schools: The real crisis in education*. New York, NY: Teachers College Press.
- Brophy, T. (1995). *Evaluation of music teachers: Toward defining an appropriate instrument*. Unpublished paper. Retrieved from <http://eric.ed.gov/PDFS/ED375029.pdf>
- Buckley, K., & Marion, S. (2011). *A survey of approaches used to evaluate educators in non-tested grades and subjects*. National Center for the Improvement of Educational Assessment. Washington, D.C. Retrieved from <http://www.nciea.org/publications/Summary%20of%20Approaches%20for%20non-tested%20gradesKBSM2011.pdf>
- Campbell, T.F. (2013). *Teacher supervision and evaluation: A case study of administrators' and teachers' perceptions of mini observations*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3556919)
- Center for Fine Arts Education. (2013). *RTTT performing fine arts assessment project*. Retrieved from <http://cfaefl.org/AssessmentProject/InnerPage.aspx?ID=1>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Colby, S.A., Bradshaw, L.K., & Joyner, R.L. (2002). *Perceptions of teacher evaluation systems and their impact on school improvement, professional development, and student learning*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Collins, I.H. (1996). Assessment and evaluation in music teacher education. *Arts Education Policy Review*, 98 (1), 16-21.

- Collins, J.B., & Pratt, D. D. (2011). The teaching perspectives inventory at 10 years and 100,000 respondents: Reliability and validity of a teacher self-report inventory. *Adult Education Quarterly: A Journal of Research and Theory*, 61(4), 358-375.
- Colwell, R. (2010). *Improving music teacher quality: Town hall meeting*. Rhode Island College. Retrieved from <http://www.ustream.tv/record/3837142>
- Cook, C., Heath, F., & Thompson, R.L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*, 60(6), 821-836.
- Crowe, E. (2008). Teaching as profession: A bridge too far? In M. Cochran-Smith, S. Feiman-Nemser, & J. McIntyre (Eds.), *Handbook of research on teacher education, 3rd edition* (989-999). New York, NY: Routledge.
- Csipkes, N.E. (2011). *Evaluating music teachers from two perspectives: Administrator and music educator*. Available from ProQuest Dissertations and Theses database. (UMI No. 1497105)
- Cuban, L. (2004). Looking through the rearview mirror at school accountability. In K.A. Sirotnik, (Ed.), *Holding accountability accountable: What ought to matter in public education*. New York, NY: Teachers College Press.
- Danielson, C. (2014). *General questions about the framework*. Retrieved from <http://www.danielsongroup.org/questions-about-the-framework-for-teaching/>.
- Danielson, C. & McGreal, T.L. (2000). *Teacher evaluation: To enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (1990a). Instructional policy into practice: The power of the bottom over the top. *Educational Evaluation and Policy Analysis*, 12(3), 339-347.

- Darling-Hammond, L. (1990b). Teacher professionalism: Why and how. In A. Lieberman (Ed.), *Schools as collaborative cultures: Creating the future now* (25–50). London: The Falmer Press.
- Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Education*, 4-14, 44.
- Darling-Hammond, L., Wise, A.E., & Pease, S.R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Day, C., Elliot, B., & Kington, A. (2005). Reform, standards and teacher identity: Challenges of sustaining commitment. *Teaching and Teacher Education*, 21(5), 563-577.
- Del Savio, C.M. (1992). *Teacher evaluation policy in Connecticut: The effect of the reform movement on teaching as a profession*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9228456)
- Delaney, D.W. (2007). *Elementary general music teachers' evaluation of instruction: Their own and others'*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3255116)
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, 36, 1-11.
- Demarest, E.J. (2010). *A learning-centered framework for education reform: What does it mean for national policy?* New York, NY: Teachers College Press.
- Doerksen, D.P. (2006). *Evaluating teachers of music performance groups*. Lanham, MD: Rowman & Littlefield Education.

- Doherty, K.M., & Jacobs, S. (2013). *State of the states 2013: Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. National Council on Teacher Quality. Retrieved from [http://www.nctq.org/dmsView/State\\_of\\_the\\_States\\_2013\\_Using\\_Teacher\\_Evaluations\\_NCTQ\\_Report](http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report)
- Doyle, W., & Ponder, G. (1977). The practicality ethic in teacher decision-making. *Interchange*, 8(3), 1-12.
- Duke, D.L. (1990). Developing teacher evaluation systems that promote professional growth. *Journal of Personnel Evaluation in Education*, 4(2), 131-144.
- Duke, D.L. (1995). *Teacher evaluation policy: From accountability to professional development*. Albany, NY: State University of New York Press.
- Duke, D.L., & Stiggins, R.J. (1986). *Teacher evaluation: Five keys to growth*. Washington, DC: American Association of School Administrators.
- Duke, R.A., & Blackman, M.D. (1991). The relationship between observers' recorded teacher behavior and evaluation of music instruction. *Journal of Research in Music Education*, 39(4), 290-297.
- Dworkin, A.G. (1997). Coping with reform: The intermix of teacher morale, teacher burnout, and teacher accountability. In Biddle, B.J., Good, T.L., Goodson, I. (Eds.), *International handbook of teachers and teaching* (459-498). Boston, MA: Kluwer Academic Publishers.
- Elpus, K. (2011). Merit pay and the music teacher. *Arts Education Policy Review*, 112 (4), 180-190.
- Emert, D., Sheehan, S., & Deitz, O.D. (2013). Music teacher evaluation in Pennsylvania. *Music Educators Journal*, 100(1), 30-31.

- Evans, J.R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2), 195-219.
- Evans, L. (1997). Understanding teacher morale and job satisfaction. *Teaching and Teacher Education*, 13(8), 831-845.
- Figlio, D., & Loeb, S. (2011). School accountability. In Hanushek, E.A., Machin, S., & Woessmann, L. (Eds.), *Handbooks in economics: Economics in education*, 3<sup>rd</sup> edition. The Netherlands: North-Holland.
- Finkel, L.S. (1983). Forward. In T.L. McGreal (Ed.), *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Finnegan, M.F. (2003). *The evaluation of high school band directors modeled after National Board for Professional Teaching Standards (NBPTS) teacher evaluation*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3167244)
- Firestone, W.A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, 20(10), 1-8.
- Flores, M.A. (2012). The implementation of a new policy on teacher appraisal in Portugal: How do teachers experience it at school? *Educational Assessment, Evaluation, and Accountability*, 24(4), 351-368.
- Florida Department of Education. (2012). *Educator quality update on teacher evaluation*. Retrieved from <http://www.fldoe.org/profdev/pdf/OverviewFloridasTeacherEvaluationSystem.pdf>
- Folger, R. (1987). Distributive and procedural justice in the workplace. *Social Justice Research*, 1(2), 143-159.

- Fricker, R.D., & Schonlau, M. (2002). Advantages and disadvantages of Internet surveys: Evidence from the literature. *Field Methods*, 14(4), 347-367.
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. London: Falmer Press.
- Gardner, R.D. (2010). Should I stay or should I go? Factors that influence the retention, turnover, and attrition of K–12 music teachers in the United States. *Arts Education Policy Review*, 111(3), 112-121.
- Geer, J.G. (1991). Do open-ended questions measure salient issues? *Public Opinion Quarterly*, 55(3), 360–70.
- Gerrity, K.W. (2013). Measuring music education: Teacher evaluation in Indiana. *Music Educators Journal*, 99(4), 17-19.
- Gipson, R.C. (1978). *An observational analysis of wind instrument private lessons*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 7818757)
- Goddard, H.J. (2004). *Elementary music teachers' and principals' perceptions of music teacher evaluation*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3144268)
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. National Comprehensive Center for Teacher Quality. Washington, D.C.  
Retrieved from <http://files.eric.ed.gov/fulltext/ED521228.pdf>

- Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning*. National Comprehensive Center for Teacher Quality. Washington, D.C. Retrieved from <http://files.eric.ed.gov/fulltext/ED532775.pdf>
- Goe, L., & Holdheide, L. (2011). *Measuring teachers' contributions to student learning growth for nontested grades and subjects*. National Comprehensive Center for Teacher Quality. Washinton, D.C. Retrieved from <http://www.gtlcenter.org/sites/default/files/docs/MeasuringTeachersContributions.pdf>
- Goe, L., Holdheide, L., & Miller, T. (2011). *Practical guide to designing comprehensive teacher evaluation systems*. Center on Great Teachers and Leaders at American Institutes for Research. Retrieved from <http://www.gtlcenter.org/sites/default/files/docs/practicalGuideEvalSystems.pdf>
- Goldrick, L. (2002). *Improving teacher evaluation to improve teaching quality*. National Governors' Association Center for Best Practices. Washington, D.C. Retrieved from <http://www.nga.org/files/live/sites/NGA/files/pdf/1202IMPROVINGTEACHEVAL.pdf>
- Gottlieb, D. (2013). Eisner's evaluation in the age of Race to the Top. *Curriculum and Teaching Dialogue*, 15(1-2), 11-25.
- Graham, P.A. (2005). *Schooling America: How the public schools meet the nation's changing needs*. New York, NY: Oxford University Press.
- Gratton, R. (2004). Teacher appraisal: A lesson on confusion over purpose. *The International Journal of Educational Management*, 18(5), 292-296.

- Hallinger, P., Heck, R.H., Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation, and Accountability*, 26(5), 5-27.
- Hasty, N.L. (2013). *Educators' experiences with new approaches to teacher evaluation*. (Doctoral Dissertation). Michigan State University, East Lansing.
- Heerwegh, D. (2004). *Using progress indicators in web surveys*. Paper presented at the 59<sup>th</sup> American Association for Public Opinion Research, Phoenix, AZ.
- Heneman, R.G., & Milanowski, A.T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(2), 173-195.
- Hill, H.C. (2009). Evaluating value-added models: A validity argument approach. *Journal of Policy Analysis and Management*, 28(4), 700-709.
- Hirokawa, J.O. (2013). *Evaluating music teachers: A comparison of evaluations by observers with varied levels of musical and observational background*. Available from ProQuest Dissertations and Theses database. (UMI No.3575268)
- Hox, J.J., & de Leeuw, E.D. (1994). A comparison of nonresponse in mail, telephone, and face-to-face surveys. *Quality and Quantity*, 28, 329-344.
- Jones, S.R., Torres, V., & Arminio, J. (2006). *Negotiating the complexities of qualitative research in higher education: Fundamental elements and issues*. New York, NY: Routledge.
- Kaestle, C.F., & Lodewick, A.E. (2007). *To educate a nation: Federal and national strategies of school reform*. Lawrence, KS: University Press of Kansas.

- Kaplowitz, M. D., Hadlock, T. D., Levine, R. (2004). A comparison of Web and mail survey response rates. *Public Opinion Quarterly*, 68(1), 94-101.
- Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology*, 85(5), 708-723.
- Kelchtermans, G. (2005). Teachers' emotions in educational reforms: Self-understanding, vulnerable commitment, and micropolitical literacy. *Teaching and Teacher Education*, 21(8), 995-1006.
- Kelly, K.O., Ang, S.Y.A., Chong, W.L., & Hu, W.S. (2008). Teacher appraisal and its outcomes in Singapore primary schools. *Journal of Educational Administration* 46(1), 39-54.
- Kennedy, M.M. (2010). Introduction: The uncertain relationship between teacher assessment and teacher quality. In M.M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (1-6). San Francisco, CA: Jossey-Bass.
- Kimball, S.M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241-268.
- Konovsky, M.A., Folger, R., & Cropanzano, R. (1987). Relative effects of procedural and distributive justice on employee attitude. *Representative Research in Social Psychology*, 17(1), 15-24.
- Koziol, S. M., & Bruns, P. (1986). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *The Journal of Educational Research*, 79(4), 205-209.
- Krosnick, J.A. (1999). Survey research. *Annual Review of Psychology*, 50, 537-567.

- Kurkul, W.W. (2007). Nonverbal communication in one-to-one music performance instruction. *Psychology of Music, 35*(2), 327-362.
- Lasky, S. (2005). A sociocultural approach to understanding teacher identity, agency, and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education, 21*(8), 899-916.
- Lavigne, A.L., & Good, T.L. (2014). *Teacher and student evaluation: Moving beyond the failure of school reform*. New York, NY: Routledge.
- LeCompte, M.D., & Schensul, J.J. (1999). *Analyzing and interpreting ethnographic data*. Walnut Creek, CA: Altamira Press.
- Legette, R. M. (2003). Multicultural music education: Attitudes, values, and practices of public school music teachers. *Journal of Music Teacher Education, 13*(1), 51-59.
- Leithwood, K., & Earl, L. (2000). Educational accountability effects: An international perspective. *Peabody Journal of Education, 75*(4), 1-18.
- Levin, H.M. (1974). A conceptual framework for accountability in education. *The School Review, 82*(3), 363-391.
- Ma, Q., & McCord, M. (2008). Web survey design. In R.A. Reynolds, R. Woods, & J.D. Baker (Eds.), *Handbook of research on electronic surveys and measurements* (9-18). Hershey, PA: Idea Group.
- Madaus, G., Russell, M., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. Charlotte, NC: Information Age Publishing.
- Maranzano, C. (2000). Music teacher performance evaluation: a call for more inclusive models. *Journal of Personnel Evaluation in Education, 14*(3), 267-274.

- Maranzano, C. (2002). *Evaluating music teachers in Virginia: Practices and perceptions*. (Doctoral dissertation). Available from ProQuest Theses and Dissertations database. (UMI No. 3041366)
- Marzano, R.J. (2012). Teacher evaluation: What's fair? What's effective? *Educational Leadership*, 70(3), 14-19.
- Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, 21(1), 29-45.
- McAllister, L. (2008). Evaluating teaching effectiveness in music. *American Music Teacher*, 58(3), 14-17.
- McGreal, T.L. (1983). *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Miksza, P., Roeder, M., & Biggs, D. (2012). Surveying Colorado band directors' opinions of skills and characteristics important to successful music teaching. *Journal of Research in Music Education*, 57(4), 364-381.
- Milanowski, A.T., & Heneman, H.G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193-212.
- Miles, M., & Huberman, M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.
- Mohlman, G., Coladarci, T., & Gage, N. L. (1982). Comprehension and attitude as predictors of implementation of teacher training. *Journal of Teacher Education*, 33(1), 31-36.
- Morgan, G.A., Leech, N.L., Gloeckner, G.W., & Barrett, K.C. (2012). *IBM SPSS for introductory statistics: Use and interpretation* (5<sup>th</sup> Ed.). New York, NY: Routledge.
- Morrel-Samuels, P. (2003). Web surveys' hidden hazards. *Harvard Business Review*, 81, 16-17.

- Mowday, R.T., Steers, R.M., & Porter, L.W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior, 14*, 224-247.
- Muncey, D., & McQuillan, P. (1996). *Reform and resistance in schools and classrooms: An ethnographic view of the Coalition of Essential Schools*. New Haven, CT: Yale University Press.
- O'Malley, K. J., Murphy, S., McClarty, K. L., Murphy, D., & McBride, Y. (2011). Overview of student growth models. (Test, Measurement, & Research Services White Paper). Iowa City, IA: Pearson. Retrieved from [https://www.pearsonassessments.com/hai/Images/tmrs/Student\\_Growth\\_WP\\_083111\\_FINAL.pdf](https://www.pearsonassessments.com/hai/Images/tmrs/Student_Growth_WP_083111_FINAL.pdf).
- Peterson, A.D. (2014). A view of current evaluative practices in instrumental music teacher education. *Arts Education Policy Review, 115*(4), 121-130.
- National Association for Music Education. (2011<sup>a</sup>). *Teacher evaluation: Position statement*. Retrieved from <http://musiced.nafme.org/about/position-statements/teacher-evaluation>.
- National Association for Music Education. (2011<sup>b</sup>). *NAfME recommendations for music teacher evaluation*. Retrieved from [http://advocacy.nafme.org/files/2012/04/teacher\\_evaluation\\_ad\\_guide.pdf](http://advocacy.nafme.org/files/2012/04/teacher_evaluation_ad_guide.pdf).
- National Center for Education Statistics. (2012). *Arts education in public elementary and secondary schools 1999–2000 and 2009–10*. Retrieved from <http://nces.ed.gov/pubs2012/2012014rev.pdf>.
- National Commission on Excellence in Education. (1983). *A Nation at Risk: The imperative for educational reform. A report to the Secretary of Education and the nation*. Washington, D.C.: United States Government Printing Office.

- National Council on Teacher Quality. (2014). *State-by-state evaluation timeline briefs*. Retrieved from [http://www.nctw.org/dmsStage/Evaluation\\_Timeline\\_Brief\\_Overview](http://www.nctw.org/dmsStage/Evaluation_Timeline_Brief_Overview).
- Nielsen, L.D. (2014). Teacher evaluation: Archiving teacher effectiveness. *Music Educators Journal*, 101(1), 62-69.
- Nolin, M.J., Rowand, C., & Farris, E. (1994). *Public elementary teachers' views on teacher performance evaluations*. National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubs94/94097.pdf>.
- Nulty, D.D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment and Evaluation in Higher Education*, 33(3), 301-314.
- Ohio Music Educators Association. (2012). *Ohio's teacher evaluation system*. Retrieved from [http://omea-ohio.org/Static\\_PDF/OTES/OTESOverviewforMusicTeachers.pdf](http://omea-ohio.org/Static_PDF/OTES/OTESOverviewforMusicTeachers.pdf).
- O'Malley, K.J., Murphy, S., McClarty, K.L., Murphy, D., & McBride, Y. (2011). *Overview of student growth models*. Test, Measurement and Research Services White Paper. Iowa City, IA: Pearson.
- O'Pry, S.C., & Schumacher, G. (2012). New teachers' perceptions of a standards-based performance appraisal system. *Educational Assessment, Evaluation, and Accountability*, 24(4), 325-350.
- Ornstein, A.C. (1988). The evolving accountability movement. *Peabody Journal of Education*, 65(3), 12-20.
- Orzolek, D. (2014). Measuring music education: Teacher evaluation in Minnesota. *Music Educators Journal*, 100(4), 22-24.
- Overland, C.T. (2014). Teacher evaluation and music education: Joining the national discussion. *Music Educators Journal*, 101(1), 56-62.

- Palazuelos, A.E. (2007). *Teacher control over evaluation*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3283728)
- Parkes, K. (2011). *Music teacher evaluation: Clarification and recommendations*. Position paper prepared for the Virginia Music Educators Association. Retrieved from <http://www.vamea.org/VASMTE%20Position%20Nov%202017-%20Statement%20Uniform%20Perf%20Standards%20Draft.pdf>.
- Partnership for Music Education Policy Development. (2012). *Criteria for designing a viable music teacher evaluation tool*. Retrieved from [http://pmepd.weebly.com/uploads/4/9/2/4/4924314/evaluation\\_tool\\_position\\_statement\\_dec\\_2012.pdf](http://pmepd.weebly.com/uploads/4/9/2/4/4924314/evaluation_tool_position_statement_dec_2012.pdf).
- Perie, M., & Baker, D.P. (1997). *Job satisfaction among America's teachers: Effects of workplace conditions, background characteristics, and teacher compensation*. National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubs97/97471.pdf>
- Perrine, W.M. (2013). Music teacher assessment and Race to the Top: An initiative in Florida. *Music Educators Journal*, 100(1), 39-44.
- Public Schools of North Carolina State Board of Education. (2013<sup>a</sup>). *North Carolina professional teaching standards*. Retrieved from <http://www.ncpublicschools.org/docs/effectiveness-model/ncees/standards/prof-teach-standards.pdf>.
- Public Schools of North Carolina State Board of Education. (2013<sup>b</sup>). *Analysis of student work*. Retrieved from <http://si2013.ncdpi.wikispaces.net/file/view/IHE2013.ASWDesignStudio.September2013.pdf/449968290/IHE2013.ASWDesignStudio.September2013.pdf>.
- Ravitch, D. (2002). Testing and accountability, historically considered. In W.M. Evers and H.J. Walberg (Eds.), *School accountability* (9-21). Stanford, CA: Hoover Institution Press.

- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Reault, J.A. (1998). *Professionalism and growth-oriented teacher evaluation: A cross-case study of two school districts*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9917449)
- Reform Support Network. (2012). *Measuring student growth for teachers in non-tested grades and subjects*. Retrieved from <http://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/measuring-student-growth-teachers.pdf>.
- Reja, U., Manfreda, K.L., Hlebec, V., Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Advances in methodology and statistics*, 19, 159-177.
- Rhode Island Department of Education. (2014). *Rhode Island model evaluation and support system*. Retrieved from [http://www.ride.ri.gov/Portals/0/Uploads/Documents/Teachers-and-Administrators-Excellent-Educators/Educator-Evaluation/Guidebooks-Forms/Support\\_Professionals\\_Guidebook.pdf](http://www.ride.ri.gov/Portals/0/Uploads/Documents/Teachers-and-Administrators-Excellent-Educators/Educator-Evaluation/Guidebooks-Forms/Support_Professionals_Guidebook.pdf).
- Robelen, E.W. (2013). Classroom portfolios used as alternative teacher evaluation measure. *Education Week*, 33(4). Retrieved from [http://www.edweek.org/ew/articles/2013/09/18/04arts\\_ep.h33.html](http://www.edweek.org/ew/articles/2013/09/18/04arts_ep.h33.html).
- Roberts, M.E., Steward, B.M., Tingley, D., Lucas, C. Leder-Luis, J., Gadarian, S.K., Albertson, B., & Rand, D.G. (*in press*). Structural topic models for open-ended survey responses. *American Journal of Political Science*. Retrieved from <http://scholar.harvard.edu/files/dtingley/files/topicmodelsopenendedexperiments.pdf>.
- Rockhoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252.

- Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform*. San Francisco, CA: Jossey-Bass.
- Rosenholtz, S.J. (1987). Education reform strategies: Will they increase teacher commitment? *American Journal of Education*, 95(4), 534-562.
- Ross, K.N. (2005). Sample design for educational survey research. *Quantitative research methods in educational planning*. Retrieved from [http://www.unesco.org/iiep/PDF/TR\\_Mods/Qu\\_Mod3.pdf](http://www.unesco.org/iiep/PDF/TR_Mods/Qu_Mod3.pdf).
- Saldana, J. (2012). *The coding manual for qualitative research*. Thousand Oaks, CA: Sage.
- Salkind, N. (2010). *Encyclopedia of research design*. Thousand Oaks, CA: Sage.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Schmidt, C.P. (1992). Reliability of untrained observers' evaluations of applied music instruction. *Bulletin of the Council for Research in Music Education*, 112, 17-28.
- Schmidt, M., & Datnow, A. (2005). Teachers' sense-making about comprehensive school reform: The influence of emotions. *Teaching and Teacher Education*, 21(8), 949-965.
- Schumacher, G. (2010). Complex teacher evaluation systems can produce negative perceptions. *ERS Spectrum*, 28(2), 27-36.
- Schuman, H, & Presser, S. (1996). *Questions and answers in attitude surveys: experiments on question form, wording, and context*. Thousand Oaks, CA: Sage.
- Shaw, R. (2013). *Music teacher evaluation in Michigan: A survey of practices and beliefs*. Paper presented at the biannual Symposium on Music Teacher Education, Greensboro, NC.

- Sheehan, K.B., & McMillan, S.J. (1999). Response variation in e-mail surveys: An exploration. *Journal of Advertising Research, 39*(4), 45-54.
- Shih, T., & Fan, X. (2008). Comparing response rates from Web and mail surveys: A meta-analysis. *Field Methods, 20*(3), 249-271.
- Shulman, L.S. (1987). Knowledge and teaching; Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-21.
- Sindberg, L. (2011). Alone all together: The conundrum of music teacher isolation and connectedness. *Bulletin of the Council for Research in Music Education, 189*, 7-22.
- Sindberg L., & Lipscomb, S. D., (2005). Professional isolation and the public school music teacher. *Bulletin of the Council for Research in Music Education, 166*, 43-56.
- Society for Music Teacher Education. (2013). *Teacher evaluation*. Retrieved from <http://smte.us/teacher-evaluation/>.
- Spillane, J.P., Reiser, B.J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*(3), 387-431.
- Spring, J. (2011). *The American school: A global context from the Puritans to the Obama era* (8<sup>th</sup> ed.). New York, NY: McGraw Hill.
- Stamm, M. (2013). My new grade: A music teacher's experience with a teacher evaluation pilot. *Triad, 80*(4), 60-63.
- Stiggins, R.J. (1986). Teacher evaluation: Accountability and growth systems – Different purposes. *National Association of Secondary School Principals Bulletin, 70*(490), 51-58.
- Stiggins, R.J., & Duke, D.L. (1988). *The case for commitment to teacher growth: Research on teacher evaluation*. Albany, NY: State University of New York Press

- Stronge, J. H. (1995). Balancing individual and institutional goals in educational personnel evaluation: A conceptual framework. *Studies in Educational Evaluation, 21*, 131-151.
- Stronge, J.H. (2006). *Teacher evaluation and school improvement: Improving the educational landscape*. Retrieved from [http://www.corwin.com/upm-data/7808\\_Stronge01.pdf](http://www.corwin.com/upm-data/7808_Stronge01.pdf).
- Stronge, J.H. (2010). *Evaluating what good teachers do: Eight research-based standards for assessing teacher excellence*. New York, NY: Routledge.
- Taebel, D.K. (1990a). An assessment of the classroom performance of music teachers. *Journal of Research in Music Education, 38*(1), 5-23.
- Taebel, D. K. (1990b). Is evaluation fair to music educators? *Music Educators Journal, 76*(6), 50–54.
- Tennessee Department of Education. (2013). *Tennessee Educator Accelerator Model (TEAM)*. Retrieved from <http://team-tn.org/teacher-model>.
- Tennessee Fine Arts Student Growth Measures System. (2013). *Teacher's guide 2013–2014*. Retrieved from <http://edweek.org/media/tdoe%20arts%20system%20teachers%20guide%2030.pdf>.
- Thomson, D.L. (2013). *Teachers' perceptions of the new Massachusetts teacher evaluation instrument and process on instructional practice*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3608479)
- United States Department of Education. (2010). *Race to the Top program: Guidance and frequently asked questions*. Washington, D.C. Retrieved from <http://www2.ed.gov/programs/racetothetop/faq.pdf>.

- United States Department of Education. (2011). *Great Teachers and Leaders: State considerations on building systems of educator effectiveness* (ED-ESE-10-0-0087). Fairfax, VA: ICF International.
- United States Government Accountability Office. (2013). *Race to the Top: States implementing teacher and principal evaluation systems despite challenges* (GAO-13-777). Retrieved from <http://www.gao.gov/assets/660/657936.pdf>.
- United States 89<sup>th</sup> Congress. (1965). *Elementary and Secondary Education Act of 1965*. Retrieved from <http://files.eric.ed.gov/fulltext/ED017539.pdf>.
- Visser, P.S., Krosnick, J.A., Marquette, J., & Curtin, M. (1996). Mail survey for election forecasting? An evaluation of the Columbus Dispatch poll. *The Public Opinion Quarterly*, 60(2), 181-227.
- Weisberg, D., Sexton, S. Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project. Retrieved from <http://files.eric.ed.gov/fulltext/ED515656.pdf>.
- Wiener, R., & Lundy, K. (2014). Survey says: Using teacher feedback to bolster evaluation. *American Educator*, 38(1), 14-17.
- Wise, A.E., & Darling Hammond, L. (1985). Teacher evaluation and teaching professionalism. *Educational Leadership*, 42(4), 28-33.
- Wise, A.E., Darling-Hammond, L., McLaughlin, M.W., & Bernstein, H.T. (1984). Teacher evaluation: A study of effective practices. *Teacher evaluation: A study of effective practices*. Santa Monica, CA: Rand.

Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.

Zimmerman, S., & Deckert-Pelton, M., (2003). Evaluating the evaluators: Teachers' perceptions of the principal's role in professional evaluation. *National Association of Secondary School Principals Bulletin, 87*, 28-37.

## Appendix A

### Invitation for State MEA Participation

Dear (Insert State MEA President and Executive Director Names Here),

My name is Lisa Martin, and I am currently a K–12 music teacher in Colorado. I am also working on completing my doctoral degree in music education at the University of Colorado Boulder. As such, I am engaging in a research project investigating how music teachers are evaluated in various states. In particular, I am interested in learning more about how music teachers perceive the evaluation process.

To gain a comprehensive picture of the different ways in which music teachers are evaluated across several states, Tennessee has been selected to participate in this study, along with Florida, Delaware, Rhode Island, and North Carolina, as you have all implemented new and innovative evaluation procedures in the past several years.

As part of my research, I would like to request your assistance in distributing an online questionnaire to practicing K–12 music teachers that are members of the Tennessee MEA. Your involvement would be limited to sending these members one invitation email and one follow-up reminder email within the next couple of weeks. I have attached these notices to this email for your review. I have also attached a copy of the survey instrument itself. In all, the questionnaire should take participants around 12 minutes to complete, and it is completely voluntary and anonymous.

In exchange for your assistance in conducting this study, I am willing to either provide a report summary for your state, or make a presentation on the findings at your state or regional MEA conference.

After reviewing the attached materials, if you are willing to assist distributing this questionnaire to your practicing K–12 music educators, please contact me as soon as possible so that we can move forward in a timely fashion. If you decide not to participate, please notify me, as well.

I look forward to your reply, and I sincerely hope we are able to work together on this very important project! If you have any questions, please call me at (847) xxx-xxxx or email me at [lisa.martin@colorado.edu](mailto:lisa.martin@colorado.edu). Thank you so much!

Sincerely,  
Lisa Martin

**Appendix B**  
**Introductory Email Script**

Send to all practicing K–12 music educators in your state by \_\_\_\_\_ (Insert Date) \_\_\_\_\_

Dear Fellow Music Educator,

I hope your school year is going well!

With the support of the \_\_\_\_\_ (Insert State Name) \_\_\_\_\_ Music Educators Association, I am writing to invite you to participate in a brief study on music teachers' impressions of teacher evaluation practices. As a fellow K–12 music teacher and doctoral candidate in music education, I'm interested in how teacher evaluation is playing out for music educators in other states.

Your involvement would entail you completing a brief, 12-minute online questionnaire. You're invited to participate in this online survey because you are currently teaching music in a state that has developed a new teacher evaluation system in the past three years.

Your participation in this study is completely voluntary and anonymous. If you choose to participate the link below will direct you toward all relevant study information, as well as the questionnaire itself. The study will be open for approximately two weeks. You will have until 11:59pm on Friday, March 14<sup>th</sup>, 2014 to complete and submit the questionnaire.

<http://tinyurl.com/MusicTeacherEvaluation>

I really do appreciate your consideration, as I know how busy life can get. If you have any questions or concerns, please contact me via email at [lisa.martin@colorado.edu](mailto:lisa.martin@colorado.edu). I hope you have a terrific rest of the school year!

All the best,  
Lisa Martin  
[lisa.martin@colorado.edu](mailto:lisa.martin@colorado.edu)  
IRB Protocol# 14-0029

## Appendix C

### Reminder Email Script

Send to all practicing K–12 music educators in your state on \_\_\_\_\_ (Insert Date) \_\_\_\_\_

Dear Fellow Music Educator,

Recently, you received an email inviting you to participate in a brief online survey regarding music teacher evaluation practices in your state. You have been asked to participate in this online survey because you are currently a music educator in \_\_\_\_\_ (Insert State Here) \_\_\_\_\_, which has implemented a new teacher evaluation system in the past three years.

If you have not yet completed this 12-minute online questionnaire and you are still interested in doing so, I hope to collect all responses by Friday, March 14<sup>th</sup>, 2014.

Please remember, completing this questionnaire is completely voluntary and anonymous. If you are interested in finding out more about this study, please click on the link below. It will direct you to more detailed information about the online survey, as well as the questionnaire itself.

**<http://tinyurl.com/MusicTeacherEvaluation>**

Thank you again for your consideration. As a fellow music educator, I know your time is valuable!

All the best,  
Lisa Martin  
lisa.martin@colorado.edu  
IRB Protocol# 14-0029

## Appendix D

### Music Teacher Evaluation Inventory (MTEI)

The following items are designed to gather information regarding the formal evaluation process you undergo as a teacher in your school. For the purposes of these items, the term "evaluation" refers to the entire process by which your competency as a teacher is assessed, including - but not limited to - factors such as classroom observation, student achievement data, and your involvement in professional development opportunities.

1. When was your MOST RECENT formal teaching evaluation conducted? If you are currently undergoing a formal teaching evaluation, please mark "13-14 School Year."

- '13-'14 School Year
- '12-'13 School Year
- Prior to the '12-'13 School Year

2. Listed below are several measures of teaching often considered during the evaluation process. For your MOST RECENT teaching evaluation, indicate the extent to which each of the following measures were considered by your evaluator(s).

	I am not sure if this was considered	Not at all considered/Does not apply	Slightly considered	Considered to a moderate extent	Considered to a great extent
Formal classroom observations (planned in advance)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Informal classroom observations (drop-in)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student achievement data (growth)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional development activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student input/feedback	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Portfolio of your work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student performances/contest ratings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please list any other measures of teaching that are considered in your evaluation:

	This is considered slightly	This is considered to a moderate extent	This is considered to a great extent
Other (please specify): ___	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please specify): ___	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Indicate the type of student achievement data considered in your teaching evaluation (check all that apply).

- Non-music standardized test score data for students I do NOT teach
- Non-music standardized test score data for those students I DO teach
- State or district developed PAPER-AND-PENCIL test designed to measure music student achievement
- State or district developed PERFORMANCE TASK designed to measure music student achievement
- Portfolio-based system examining samples of my music students' work
- Other (please specify): \_\_\_\_\_

3. In item #2, you indicated which measures ARE used in your evaluations. Now, please indicate to what extent you believe each of the following measures SHOULD BE considered in the evaluation of your teaching performance.

	This should not be considered at all/Does not apply	This should be considered slightly	This should be considered to a moderate extent	This should be considered to a great extent
Formal classroom observations (planned in advance)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Informal classroom observations (drop-in)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student achievement data (growth)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional development activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student input/feedback	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Portfolio of your work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student performances/contest ratings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please list any other measures you feel should be considered in your evaluation.

	This should be considered slightly	This should be considered to a moderate extent	This should be considered to a great extent
Other (please specify): ____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please specify): ____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. For your MOST RECENT evaluation, which of the following people provided input on your overall evaluation? (check all that apply)

- Building administrator (Principal/Assistant Principal) with NO music teaching expertise
- Building administrator (Principal/Assistant Principal) WITH music teaching expertise
- District music or fine arts administrator (Supervisor/Coordinator)
- Other music teacher(s)
- Other teacher(s) (non-music)
- Other (please specify): \_\_\_\_\_

5. Of those you listed in response to item #5, who has the MOST AUTHORITY in determining your overall evaluation? (select one)

- Building administrator (Principal/Assistant Principal) with NO music teaching expertise
- Building administrator (Principal/Assistant Principal) WITH music teaching expertise
- District music or fine arts administrator (Supervisor/Coordinator)
- Other music teacher(s)
- Other teacher(s) (non-music)
- Other (please specify): \_\_\_\_\_

6. Indicate the extent to which you agree or disagree with each statement by marking the appropriate box. For the purposes of the following items, the term "evaluator" refers to the individual who has the MOST AUTHORITY in determining your overall evaluation (i.e., reflecting your answer to #5).

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
My evaluator does a thorough job.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My evaluator has considerable experience in teaching.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I trust my evaluator to accurately assess my teaching performance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have a good interpersonal relationship with my evaluator.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My evaluator is qualified to assess my teaching.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My evaluator is fair and objective.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please describe any other factors about your primary evaluator that you believe are relevant: \_\_\_\_\_

7. Upon completion of your MOST RECENT evaluation, you received or expect to receive (check all that apply):

- An evaluation summary in written form
- A conference with your primary evaluator to review your evaluation
- A brief verbal report
- None of the above
- Other (please specify): \_\_\_\_\_

8. Teacher evaluations can fulfill many objectives. Indicate the degree to which each objective IS CONSIDERED in your school's teacher evaluation process.

	I am not sure if this is considered an objective	Not considered an objective at all	To a small extent, this is an objective	To a moderate extent, this is an objective	To a great extent, this is an objective
To guide improvement of teachers' skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To discharge incompetent teachers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To determine teachers' merit pay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To help teachers better focus upon student outcomes and growth	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To recognize and reinforce excellent teaching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To make tenure and promotion decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please describe any other objectives of your current teacher evaluation process:

	To a small extent, this is an objective	To a moderate extent, this is an objective	To a great extent, this is an objective
Other (please specify):__	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please specify):__	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. In item #8, you identified the objectives considered in your evaluation. For this item, indicate the degree to which each objective SHOULD BE CONSIDERED in your school's teacher evaluation process.

	Should not be considered an objective at all	To a small extent, this should be considered an objective	To a moderate extent, this should be considered an objective	To a great extent, this should be considered an objective
To guide improvement of teachers' skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To discharge incompetent teachers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To determine teachers' merit pay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To help teachers better focus upon student outcomes and growth	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To recognize and reinforce excellent teaching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To make tenure and promotion decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please describe any other objectives you believe SHOULD BE CONSIDERED in your teacher evaluation process.

	To a small extent, this should be an objective	To a moderate extent, this should be an objective	To a great extent, this should be an objective
Other (please specify): ____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please specify): ____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Items in the previous section focused on characteristics of the evaluation process. Items on this page will address your perceptions of the evaluation process.

10. Consider your overall impressions of the teacher evaluation process used in your school/district. Indicate the extent to which you agree or disagree with each statement below.

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
I would choose to participate in our teacher evaluation process even if it was not required.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Our teacher evaluation process is clear to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am satisfied with our teacher evaluation process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Our teacher evaluation process is not worth the effort.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Our teacher evaluation process equitably assesses my performance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Our evaluation process provides an accurate depiction of my teaching performance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My school should continue to implement the current evaluation process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11. Indicate the extent to which you agree or disagree with each statement concerning your teacher evaluation process:

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
The evaluation process has helped me become a more effective teacher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The achievement of my students has improved as a result of our evaluation process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have used feedback from the evaluation process to help identify relevant professional development opportunities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The evaluation process motivates me to improve my teaching practice.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The evaluation process causes me to reflect more upon my teaching.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The evaluation process contributes to improved student learning.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

12. Indicate the nature and extent to which you believe the teacher evaluation process impacts each aspect of your teaching:

	Extremely Negative Impact	Negative Impact	No Impact	Positive Impact	Extremely Positive Impact
My ability to make independent choices regarding daily classroom practice.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to implement appropriate instructional methodologies.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to utilize planning and preparation time in a manner that best suits students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to assess students in a manner relevant to my content area.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to select appropriate instructional materials.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to utilize instructional time in a manner that best benefits students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

13. Describe how the teacher evaluation process at your school impacts music student learning and/or your effectiveness as a music teacher.

14. How has the teacher evaluation process impacted your perception of music teaching as a profession? Does this process elevate or diminish the professional status of music teachers?

The following items are designed to gather information about you and your teaching experience.

15. In what state do you currently teach?

- Delaware
- Florida
- North Carolina
- Rhode Island
- Tennessee

16. How many total years of K-12 music teaching experience do you have?

17. Do you currently have tenure/non-probationary status?

- Yes
- No

18. How many more years do you expect to remain a K-12 music teacher?

You're almost done!

19. Indicate the type of school in which you currently teach.

- Public School
- Charter School
- Private School
- Other (please specify): \_\_\_\_\_

20. Please indicate the level in which you spend the MAJORITY of your time teaching.

- Elementary School (K-5)
- Middle School (6-8)
- High School (9-12)
- Other (please specify): \_\_\_\_\_

21. Please indicate the area in which you spend the MAJORITY of your time teaching.

- General Music
- Band
- Orchestra
- Vocal Music
- Other (please specify): \_\_\_\_\_

22. Is 50% or more of your teaching load dedicated to music classes?

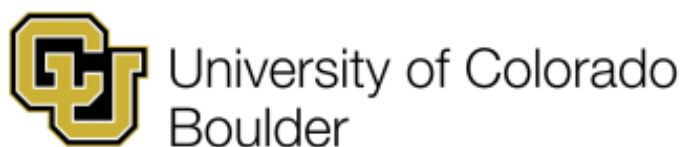
- Yes
- No

23. What is your gender?

- Male
- Female
- Other/prefer not to respond

## Appendix E

### Institutional Review Board (IRB) Approval



Institutional Review Board  
563 UCB  
Boulder, CO 80309  
Phone: 303.735.3702  
Fax: 303.735.5185  
FWA: 00003492

28-Jan-2014

### Exempt Certification

Martin, Lisa  
**Protocol #:** 14-0029  
**Title:** Music Educators' Impressions of Teacher Evaluation Practices

Dear Lisa Martin,

The Institutional Review Board (IRB) has reviewed this protocol and determined it to be of exempt status in accordance with Federal Regulations 45 CFR 46.101(b). Principal Investigators are responsible for informing the IRB of any changes or unexpected events regarding the project that could impact the exemption status. Upon completion of the study, you must submit a Final Review via eRA. It is your responsibility to notify the IRB **prior** to implementing any changes.

**Certification Date:** 28-Jan-2014  
**Exempt Category:** 2

Click here to find the IRB reviewed documents for this protocol: [Study Documents](#)

The IRB has reviewed this protocol in accordance with federal regulations, university policies and ethical standards for the protection of human subjects. In accordance with federal regulation at 45 CFR 46.112, research that has been approved by the IRB may be subject to further appropriate review and approval or disapproval by officials of the institution. The investigator is responsible for knowing and complying with all applicable research regulations and policies including, but not limited to, Environmental Health and Safety, Scientific Advisory and Review Committee, Clinical and Translational Research Center, and Wardenburg Health Center and Pharmacy policies.

Please contact the IRB office at 303-735-3702 if you have any questions about this letter or about IRB procedures.

Douglas Grafel  
IRB Admin Review Coordinator  
Institutional Review Board

## Appendix F

### Pilot Version of Music Teacher Evaluation Inventory

#### Evaluation Procedures

The following items are designed to gather information regarding the formal evaluation process you undergo as a teacher in your school. For the purposes of these items, the term "evaluation" refers to the entire process by which your competency as a teacher is assessed, including – but not limited to – factors such as classroom observation, student achievement data, and your involvement in professional development opportunities.

1. When was your MOST RECENT formal teaching evaluation conducted? If you are currently undergoing a formal teaching evaluation, please mark "13-14 School Year."

- '13-'14 School Year  
 '12-'13 School Year  
 '11-'12 School Year  
 '10-'11 School Year  
 Prior to the '10-'11 School Year  
 I do not know

Other (please specify)

2. Listed below are several measures of teaching often considered during the evaluation process. For your MOST RECENT teaching evaluation, indicate the extent to which each of the following measures were considered.

	I'm not sure if this was considered	Not at all considered	Slightly considered	Considered to a moderate extent	Considered to a great extent
Formal classroom observations (planned in advance)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Informal classroom observations (drop-in)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student achievement data (student growth)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional development activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student survey data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Portfolio of your work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student performances/contest ratings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

3. Indicate the type of student achievement data considered in your teaching evaluation (check all that apply).

- Non-music standardized tests score data for students I do NOT teach
- Non-music standardized test score data for those students I DO teach
- State or district developed PAPER-AND-PEN®IL t st designed t measure music student achievement
- State or district developed PERFORMANCE TASK designed to measure music student achievement
- Portfolio-based system examining samples of my music students' work

Other (please specify)

4. Please indicate to what extent you believe each of the following measures SHOULD BE considered in the evaluation of your teaching performance.

	This should not be considered at all	This should be considered slightly	This should be considered to a moderate extent	This should be considered to a great extent
Formal classroom observations (planned in advance)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Informal classroom observations (drop-in)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student achievement data (student growth)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional development activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student survey data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Portfolio of your work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student performances/contest ratings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

5. For your MOST RECENT evaluation, which of the following provided input on your overall evaluation? (check all that apply)

- Building administrator (Principal/Assistant Principal) with NO music expertise
- Building administrator (Principal/Assistant Principal) WITH music expertise
- District music or fine arts administrator (Supervisor/Coordinator)
- Other music teacher(s)
- Other teacher(s) (non-music)
- Students
- Parents

Other (please specify)

6. Of those you listed in response to item #5, who has the MOST AUTHORITY in determining your overall evaluation?

- Building administrator (Principal/Assistant Principal) with NO music expertise
- Building administrator (Principal/Assistant Principal) WITH music expertise
- District music or fine arts administrator (Supervisor/Coordinator)
- Other music teacher(s)
- Other teacher(s) (non-music)
- Students
- Parents

Other (please specify)

7. Read each statement. Indicate the extent to which you agree or disagree with each statement by marking the appropriate box. For the purposes of the following items, the term "evaluator" refers to the individual who has the most authority in determining your overall evaluation (reflecting your answer to #6).

	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
My evaluator does a thorough job.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My evaluator has considerable experience in teaching.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I trust my evaluator to accurately assess my teaching performance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have a good interpersonal relationship with my evaluator.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My evaluator is qualified to assess my teaching.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My evaluator is fair and objective.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please describe any other factors about your primary evaluator that you believe are relevant:

8. Upon completion of your MOST RECENT evaluation, you received (check all that apply):

- An evaluation summary in written form
- A conference with your primary evaluator to review your evaluation
- A brief verbal report
- None of the above

Other (please specify)

9. Teacher evaluations can have a variety of objectives. For example, teacher evaluations might be used to inform tenure and promotion decisions. Read the list of objectives, and indicate the degree to which each IS considered in your school's current teacher evaluation process.

	I am not sure if this is considered an objective	Not considered an objective at all	To a small extent, this is an objective	To a moderate extent, this is an objective	To a great extent, this is an objective
To guide improvement of teachers' skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To better plan in-service professional development opportunities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To discharge incompetent teachers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To determine teachers' pay levels	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To help teachers better focus upon student outcomes and growth	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To recognize and reinforce excellent teaching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To make tenure and promotion decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

10. Read the list of objectives, and indicate the degree to which you believe each SHOULD BE considered in your teacher evaluation process.

	Should not be considered an objective at all	To a small extent, this is should be an objective	To a moderate extent, this should be an objective	To a great extent, this should be considered an objective
To guide improvement of teaching skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To better plan in-service professional development opportunities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To discharge incompetent teachers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To determine teachers' pay levels	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To help teachers better focus upon student outcomes and growth	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To recognize and reinforce excellent teaching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To make tenure and promotion decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

## Perceptions of Teacher Evaluation

The previous section dealt with components comprising your evaluation process. Items in this section will address your perceptions of the evaluation process.

11. Consider your overall impressions of the teacher evaluation process used in your school/district. Indicate the extent to which you agree or disagree with each statement below.

	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
I would choose to participate in our teacher evaluation process even if it wasn't required.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Our teacher evaluation process is worth the effort I must put into it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The process through which I am evaluated is clear to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am satisfied with the evaluation process by which I am held accountable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Our teacher evaluation process fairly assesses my effectiveness.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The overall rating I receive in the evaluation process is an accurate reflection of my teaching performance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The measures by which I am evaluated are clear to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My school should continue to implement the current evaluation process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The areas on which I am evaluated are within the scope of my control.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The evaluation process causes me stress.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Perceived Impact of Teacher Evaluation Procedures

Items in the previous section addressed your overall impressions of the teacher evaluation process used in your school/district. In this section, you will respond to items regarding the impact the teacher evaluation process has on various aspects of your teaching.

12. Indicate the extent to which you agree or disagree the evaluation process has impacted each scenario described below.

	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
The evaluation process has helped me become a more effective teacher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have improved my teaching practice as a result of participating in our evaluation process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The achievement of my students has improved as a result of my participation in our evaluation process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have used feedback from the evaluation process to help identify relevant professional development opportunities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The evaluation process motivates me to improve my teaching practice.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
As a result of participating in the evaluation process, I spend a greater amount of time reflecting on my teaching practice.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The evaluation process contributes to improved student learning.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The evaluation process is valuable to me in my professional development.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My students are better off as a result of my participation in the evaluation process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

13. Consider the various components of your daily teaching practice, described below. Indicate the nature and extent to which you believe the teacher evaluation process impacts each teaching component.

	Extremely Negative Impact	Negative Impact	No Impact	Positive Impact	Extremely Positive Impact
My ability to make independent choices regarding daily classroom practice.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The instructional methodologies I implement in my classroom.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The instructional materials I use in my teaching.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to assess students in a manner relevant to my content area.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to utilize planning and preparation time in a manner that best benefits students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My ability to utilize instructional time in a manner that best benefits students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14. Many teacher evaluation processes are designed to develop more effective teachers. One manner in which many gauge effective teaching is through increased student learning. Considering your own experiences, do you feel the teacher evaluation process at your school helps develop your effectiveness as a music teacher? Why or why not?

15. The concept of "quality control" is embedded in many professions. People such as doctors, lawyers, and engineers are held to certain performance standards as part of their profession. Consider music teaching as a profession, and reflect on whether you feel the teacher evaluation process is an effective means of "quality control." Does the teacher evaluation process make you feel like more of a professional, or not? Why?

## Demographic Information

The following items are designed to gather information about you and your teaching experience.

16. In what state do you currently teach?

- Delaware  
 Florida  
 North Carolina  
 Tennessee  
 Rhode Island

17. How many total years of K-12 music teaching experience do you have?

18. How many years have you taught music in your current state?

19. How many years have you been in your current teaching position?

20. How many more years do you expect to remain a K-12 music teacher?

21. Indicate the type of school in which you currently teach.

- Public school  
 Charter school  
 Private school

Other (please specify)

22. Please indicate the level in which you spend the MAJORITY of your time teaching.

- Elementary  
 Middle School  
 High School

Other (please specify)

23. Please indicate the area in which you spend the MAJORITY of your time teaching.

General Music

Band

Orchestra

Vocal Music

Other (please specify)

24. What portion of your total teaching load is represented by teaching MUSIC classes?

25. What is your gender?

Female

Male

Prefer Not To Respond

## Appendix G

### Codes, Themes, Examples, and Frequencies for Evaluator Attribute Item

	Code	Theme	Example 1	Example 2	Frequency	% of Total Comments
<b>Concerns</b>	CoCo	Concerns about Evaluation Consistency	Over the six years of teaching, I have had four evaluators along with five different principals. Consistency, or even having an evaluation, is rare.	The evaluations are not consistent from administrator to administrator.	13	13%
	CoNoMu	Concerns about Evaluator's Lack of Music Expertise	Music evaluation is foreign to the evaluator.	My evaluator cannot understand my lessons, and therefore cannot evaluate how effective I actually am in the music/band classroom.	38	39%
	CoObj	Concerns about Objectivity and Politics	[My primary evaluator] had a bad music experience in school where her parents forced her to take music and now she is extremely critical what I teach.	[My primary evaluator] allows personal differences to get in the way of his objective evaluation.	9	9%
	CoPerf	Evaluator Only Concerned About Performances	My evaluator only cares a about a great Winter Holiday Show.	As long as the concert sounds good, they leave me alone.	2	2%
<b>Attributes</b>	AttEff	Reassured by/Appreciate Evaluator's Effort	[My] primary evaluator recognizes [his] lack of expertise in my curricular area and attempts to compensate.	My evaluator does not know much about music education, but he does ask great questions that help him understand my content.	8	8%
	AttMisc	Miscellaneous Evaluator Attributes	My principal is a buffoon.	She is new to the school and a new principal.	25	26%
	AttMusic	Evaluator has Music Experience	I am very thankful that my evaluator has expertise in music.	[My evaluator] is also a classroom music teacher and supervisor.	2	2%


## Appendix H

### Codebook Development

In discussing coding labels with the expert auditor, we determined that several initial codes could be logically condensed into fewer codes representing broader conceptualizations of the emergent themes. For example, the initial codebook included one code that represented the status of music educators/education being diminished as a result of the teacher evaluation process, while a second code referenced a notion that music, as a subject, did not matter to others. Ultimately, during a subsequent review of participants' responses, it became clear that feelings of "music not mattering" in the evaluation process were indeed a reflection of the diminished status of music educators/education. Therefore, these responses were condensed into a single code, "StaDimMu" (status of music educators/education diminished).

#### *Relationship Between Initial Codes Aggregated into StaDimMu*

Code	Theme	Example
StaDimMu	Status of Music Educators/ Education Diminished	The evaluation process seriously diminishes the status of music educators.
MuNoMat	Music Does Not Matter to Others	We are being told, in a very subtle way, that we are no longer necessary, relevant, or wanted.



In terms of participants' concerns regarding the accuracy, appropriateness, and relevance of teacher evaluation in the context of music education, I began with three separate codes, summarized in Table 4.20. These three codes each addressed slightly different interpretations of

the teacher evaluation process and its relevance to music educators. However, these codes still demonstrated some overlap. To better differentiate participants' concerns, I applied the CoFITSco and CoFITGro codes implemented in the first open-ended item. This enabled me to distinguish concerns about the limited scope and applicability of the overall evaluative instrument (CoFITSco) from concerns about the validity of the test data used to evaluate music teachers (CoFITGro). I then recoded all items according to these two codes.

*Initial Codes Addressing Participants' Concerns with Accuracy, Fit, and Relevance of Teacher Evaluation in the Context of Music Teaching*

	<b>Code</b>	<b>Theme</b>	<b>Example</b>
<div style="border: 1px solid black; padding: 5px; width: fit-content;"> <p><b>CoFITSco:</b> Concerns with Fit – Limited/ Unrealistic Scope</p> </div> <div style="border: 1px solid black; padding: 5px; width: fit-content; margin-top: 5px;"> <p><b>CoFITGro:</b> Concerns with Fit – Validity of Growth Data</p> </div>	NoAccRef	Evaluation not an Accurate Reflection of Work as a Music Teacher	The main problem is when they try and correlate music classes to scores on reading tests or 7th grade English tests.
	CoFFC	Concern over Evaluation Fit/Fairness/Clarity to Music Teachers	The process is not geared towards our subject area and circumstances as a music specialist.
	IrrTask	Evaluation Includes Irrelevant Tasks	There are requirements and expectations that keep me from really teaching music.

Another initial code reflected participants' desire to leave music education (LeaveMusEd). This code was first intended to capture participants' explicit feelings about leaving the profession due to the evaluation process. As one participant wrote, "Honestly, [evaluation] is one of several factors I feel that are spurring me to change professions," while another wrote, "I want out of this profession ASAP." However, there were additional meaningful

participant contributions that reflected music educators' concerns for others choosing teaching as a profession. For example, one participant explained:

I no longer encourage my high school students to go into education. I almost tell them to run the opposite way! Something is wrong when you can't encourage someone to go into your own profession.

During the second round of coding, I resolved the inward and outward orientations to professional commitment/encouragement by creating a new, broader code, "NoTeach," which aimed to capture the feelings of those planning to leave the profession as well as those discouraging others from joining the profession.

Finally, in the initial codebook for the second open-ended item, I included a code that captured participants' sentiment that teacher evaluation did not diminish the status of music educators/education. This code, "NoDimMu," appeared to capture the same idea as the initial code "NiGE," which reflected participants' feelings regarding evaluation's lack of influence upon music education's status as a profession. To streamline the codebook, I condensed these two codes into the single theme "NiGE."