

© 2016

Domecq Smith

ALL RIGHTS RESERVED

MUSIC TEACHER EVALUATION: NEW JERSEY MUSIC TEACHER ATTITUDES
AND PERCEPTIONS OF EVALUATOR TYPES AND EVALUATION MODEL
TYPES AND THEIR INTERACTION

By

DOMECQ SMITH

A dissertation presented to the
Graduate Committee in the Mason Gross School of the Arts
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Musical Arts

Graduate Program in Music Education

Written under the direction of

William L. Berz

And approved by

William L. Berz, Advisor, Professor of Music

Stephanie Cronenberg, Assistant Professor of Music

Christopher Doll, Associate Professor of Music

Charles Maranzano, Music Education

Reginald Luke, Professor of Mathematics

Karina Bruk, Coordinator of Graduate Studies

New Brunswick, New Jersey

January, 2017

ABSTRACT OF THE DISSERTATION

Music Teacher Evaluation:

New Jersey Music Teacher Attitudes and Perceptions of Evaluator Types and Evaluation

Model Types and Their Interaction

By DOMEQC SMITH

Dissertation Director:

William L. Berz

This study examined New Jersey public school music teacher attitudes and perceptions of state mandated evaluations performed upon them by comparing teachers evaluated by administrators having a music degree with those evaluated by administrators who did not. Music teachers evaluated under two contrasting state approved evaluation models, Charlotte Danielson's Framework for Teaching, and James Stronge's Teacher Effectiveness Performance Evaluation System were included in the study. Research questions were to what extent evaluator background and evaluation model type act as independent factors or interact in influencing music teacher attitudes and perceptions of the evaluation process. The design was a 2 X 2 analysis of variance with the factors Administrator Type and Model Type. Qualitative analysis of an open-ended response was included. Participants were drawn from New Jersey K-12 active public school music teachers ($N = 308$). Outcome variables were derived from a Likert scale survey developed for this study measuring teacher attitudes and perceptions encompassing four domains. Results indicated more favorable perception between respondents when the evaluation

was performed by administrators who had a music degree than when they did not at significant ($p < .05$) and highly significant ($p < .01$) levels. Results indicated more favorable perception between respondents when the evaluation utilized the Danielson model than the Stronge model at significant ($p < .05$) levels, yet interaction between the factors of evaluator and model type were encountered at significant levels ($p < .05$) for questions pertaining to professional growth and evaluative philosophy. Findings suggest that the factors of administrator and model type did not act independently in influencing music teacher perception but operated together. Additional research directed towards administrator/model type interaction in the influence of music teacher attitude and perception is recommended.

Acknowledgements

Regarding the influence people known and unknown have upon the lives of others, someone once said, “Every person is a bus in which their ancestors are riding.”

The same can be said for a dissertation or any other major endeavor in which the bus is the work, and its passengers are those who contributed to its creation, this dissertation being no exception.

And the bus is full.

To my advisor Dr. William Berz is thanks given for his knowledge and wisdom during this journey as well as his meticulous and expert care in considering every textual element of this study, no stone left unturned. A more diffuse narrative would have resulted without his hand.

To the members of my dissertation committee, Dr. Berz, Dr. Karina Bruk, Dr. Stephanie Cronenberg, Dr. Christopher Doll, Dr. Rhonda Hackworth, and Dr. Charles Maranzano are all owed gratitude. It is an honor that these individuals were members of my committee. Particular thanks goes to committee member Dr. Maranzano, author of the first dissertation on the topic of music teacher evaluation. I thank him for his participation and valuable advice.

Enthusiastic mention goes to Dr. Reginald Luke, Chair of Mathematics of Middlesex County College, New Jersey and visiting Rutgers professor for his statistical expertise, good humor, and commitment to the project. His generosity and contribution cannot be overestimated. Thanks statistically also goes to Miguel Martinez for his suggestions.

I thank Dr. Drew Gitomer, Rose and Nicholas DeMarzo Chair in Education, Rutgers for his observations in the alerting of potential confounds during the study's formation. I also thank Dr. Francesca Murphy, Professor of Theology of the University of Notre Dame who helped to better clarify some philosophical language. I appreciate her ongoing interest in this work.

To F.W. Cook School's administrators since 2006 are given acknowledgment for their contributions ranging from administrative recommendations to the on-time arrival of my coursework. These administrators include BJ BrownJohnson, Dr. Caryn Cooper, Frank Dincuff, Dr. Laura Fatal, Dr. Christopher Lommerin, John Martucci, and Doris Williams, with fond mention of Pauline Jazikoff, former and highly esteemed secretary of Cook School.

Special thanks go to my dear friends Dr. Emmet and Judy Smith of Ft. Worth, Texas. Their friendship and support helped to make the completion of this study possible.

Thanks go to William McDevitt, president of the New Jersey Music Educators Association for agreeing to sponsor the survey through their website and also for allowing the distribution of the survey during NJMEA's February 2016 convention. A heap of gratitude goes to NJMEA's membership for their participation in this study. Betsy Maliszewsky, NJMEA presenter and fellow doctoral student is given thanks for her assistances.

In my need to better understand the characteristics of the Stronge evaluation model, thanks go to Dr. Ginny Tonneson, Vice President of Stronge and Associates for her time and clarifications.

Dr. Gayle Carrick, retired superintendent of the Cranford Public Schools, New Jersey was of invaluable help in providing information and advice regarding the mailed version of the study's survey. Her suggestions also aided in obtaining an adequate survey response rate. I am very grateful for her help and interest.

Dr. Vicky Robinson who likewise has gone down the dissertation path deserves mention for her support. Thanks also go to Albert and Therese Teglash, and Adrian and Barbara Zapotocky of Rahway, New Jersey for their contributions.

To Pastor George Faour, the clergy and people of St. John Church, Orange, New Jersey, are given thanks.

Any study dealing with quantitative analysis may need the use of Microsoft Excel at a level of ability that may be greater than average. Thanks go to Henry Hom for sharing his expertise in this area.

My family has been in the front seat with me for every mile of this journey and has contributed in so many and untold ways that space here is inadequate. I thank my children AnnaLotte, Franz, and Friedrich Smith for patiently listening to their father discuss coursework and research during the course of much of their childhood. They each deserve a medal. To Franz and my wife Peach who enthusiastically assisted in the hand-distribution and collection of the survey are given high fives each. To my wife whose intelligence and insight provided a ready and clear mirror to my thoughts every step of the way deserves more than words will express. When the bus ran out of gas, she pushed.

These words cannot end without acknowledgement of the profound influence and contribution of those who though not physically occupying a place on the bus are still

riding none the less. They are Breon and Billie Smith, parents; Dion Smith, brother; and Fr. George M. Acker, priest.

And a host of others.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	viii
List of Tables	xii
List of Figures	xiii
Chapter 1—Introduction	1
Questions for Research	6
Limitations	6
Chapter 2—Review of Literature.....	9
Foundations of Teacher Evaluation	9
Quantitative methodology.....	11
Qualitative methodology.....	12
Music Teacher Evaluation	14
Four Elements of Music Teacher Evaluation.....	15
Evaluator	26
Administrator suitability	28
Self-evaluation	37
Student/parent/peer surveys	39
Summary of Evaluator	40
Evaluation	40
Instrument suitability	41
Summary of Evaluation	48

Result	48
Professional development	49
Professional concern and perception	54
Teacher compensation	56
Summary of Result	58
Teacher Evaluation Models in New Jersey	59
Overview	59
Danielson Framework for Teaching	63
Stronge Teacher Effectiveness Performance Evaluation System	69
Other Teaching Frameworks	76
Chapter 3—Methodology	79
Introduction	79
Population	80
Instrumentation	81
Data collection	83
Data Analysis	86
Quantitative considerations	86
Qualitative considerations	87
Chapter 4—Results	89
Introduction	89
Survey Return Rate and Origin	90
ANOVA Findings Overview	93
ANOVA Findings by Domain and Question Items	96

Domain 1.....	96
Domain 2.....	96
Domain 3.....	101
Open-ended Response.....	105
Response examples	109
Chapter 5—Discussion	114
Introduction.....	114
Domain 1, “You, the Teacher”	114
Domain 2, “Evaluator and Evaluation Instrument”	118
Administrator Type.....	119
Model type	120
Domain 3 “Resulting Usefulness and Professional Growth”	123
Summary of Findings and Discussion	129
Areas for Further Research.	130
Bibliography	132
Appendix A—New Jersey Music Teacher Evaluation Survey	140
Appendix B—Informed Consent Form	144
Appendix C—New Jersey Music Teacher Evaluation Survey Hard Copy	
Introductory Page.....	145
Appendix D—Distribution T-shirts for NJMEA February 2016 Convention	146
Appendix E—New Jersey Counties and Public School Districts Determined	
to Utilize the Stronge Teacher Effectiveness Performance	
Evaluation System	147

Appendix F—Demographic Data: Responses by Frequency and Percentage	149
Appendix G—Summary Statistics for Domain 1, “You, the Teacher”	162
Appendix H—Summary Statistics for Domain 2, “Evaluator and Evaluation Instrument”	164
Appendix I—Summary Statistics for Domain 3, “Resulting Usefulness and Growth”	166
Appendix J—Rutgers IRB Approval	167

List of Tables

Table 1—Themes and Representative Sources.....	16
Table 2—Elements and Related Themes	25
Table 3—Evaluation Models Approved for Use in New Jersey as of 2015	60
Table 4—Five Most Popular Teacher Evaluation Models in Jersey as of February, 2013	61
Table 5—Charlotte Danielson’s Framework for Teaching, 2013 Edition.....	65
Table 6—James Stronge’s Teacher Effectiveness Performance Evaluation System	71
Table 7—Survey Returns by Method of Return and Response Rate.....	91
Table 8—Survey Returns from Danielson and Stronge Districts by Method of Return.....	92
Table 9—Survey Returns from Danielson and Stronge Respondents by Administrator Type	92
Table 10—Overview of Results for Domain 1, “You the Teacher”	94
Table 11—Overview of Results for Domain 2, “Evaluator and Evaluation Instrument”	95
Table 12—Overview of Results for Domain 3, “Usefulness and Professional Growth”	96
Table 13—Open-ended Response Return Rate	106
Table 14—Themes by Frequency for Danielson and Stronge Models from Respondents Evaluated by an Administrator with a Degree in Music	107
Table 15—Themes by Frequency for Danielson and Stronge Models from Respondents Evaluated by an Administrator without a Degree in Music	107

List of Figures

Figure 1—Plot for “The administrator was qualified to evaluate you”	97
Figure 2—Plot for “The person conducting the evaluation makes a meaningful contribution to your musical development and growth”	98
Figure 3—Plot for “The evaluation instrument used to evaluate you was suited to your area of specialization”	99
Figure 4—Plot for “The evaluation instrument your district uses best reflects your own evaluation philosophy”	100
Figure 5—Plot for “You would welcome the administrator to evaluate you again if you had a choice”	101
Figure 6—Plot for “You found the evaluation useful towards your professional growth”	102
Figure 7—Plot for “You incorporated the suggestions from your evaluation into your future lessons/rehearsals”	103
Figure 8—Plot for “You would recommend the same administrator in the evaluation of other music teachers in their professional growth”	104
Figure 9—Plot for “You would look forward to the next evaluation by the same administrator as an opportunity for further professional growth”	105
Figure 10—Themes by Percentage for Danielson and Stronge Models from Respondents Evaluated by an Administrator with a Degree in Music	108
Figure 11—Themes by Percentage for Danielson and Stronge Models from Respondents Evaluated by an Administrator without a Degree in Music.....	108

Chapter 1

Introduction to the Study and Rationale

Calls for teacher accountability have increased ever since the 1980s, especially so in the past decade. These measures have encouraged more detailed evaluation of teachers. One such example in New Jersey has been the implementation of formalized assessment methods to gauge teacher effectiveness in the classroom.

As regulations were developed, local school districts were either given a choice of a number of state-approved models or were allowed to create their own per state requirements. These models were not developed to be content specific, but instead were intended for general use in all subjects. This type of evaluation model is commonly referred to as a generic type of model.

Two of the most widely used systems in New Jersey at the time of this research were Charlotte Danielson's Framework for Teaching (FfT) used by 60 percent of New Jersey school districts, followed by James Stronge's Teacher Effectiveness Performance Evaluations System (TEPES) used by 11 percent of New Jersey's districts (Mooney, 2013). Each system is a conception of identified areas of teaching expressed through the creation of a framework comprised of chief components followed by smaller ones. These components broadly include preparative delivery of instruction, instruction and rapport, in addition to non-instructional professional responsibilities. Although both may appear to differ only in presentation, they differ in their degree of classroom specificity.

Danielson's FfT evolved from her conception of what constituted effective classroom practice, the evaluation model coming into being later as a result for the need to evaluate a teacher's utilization of the FfT framework within instruction. Its discreet

classroom orientation directs the observer to assess classroom specific teaching behaviors as found within FFT and not normally encountered in specialized settings such as the music lesson or rehearsal.

In contrast, Stronge's TEPES appears to focus less on discreet classroom instructional behaviors. It was conceived initially as an evaluative instrument, and on the surface appears less bound by any defined, explicit, or preconceived instructional practices or settings. According to its creator James Stronge, a hallmark of the model is its flexibility; aspects of the model may be modified to meet specific school district needs. Its wording is careful to eschew any specific instructional/learning environments.

With the implementation of these and other systems for use in New Jersey schools, along with other assessment criteria, tenure regulations have been changed, a fact that has been noted in the popular press. On August 6, 2012, Governor Chris Christie signed a reform bill making it more difficult for teachers to receive tenure while making it easier for districts to remove tenure for underperforming teachers. New Jersey Senate President Stephen Sweeney referred to the signing as historic on account that it was altering a system that had been in place for over a century. "We can't have the bad ones in the schools anymore," Sweeney said. "One bad teacher is one bad teacher too many" (Renshaw, 2012). The new evaluation models were a part of this reform.

Some teachers, as well as leaders within the New Jersey Education Association (NJEA) see problems arising from these new assessment strategies. The validity that the new evaluation models possess in accurately and fairly assessing a teacher's ability is a contentious subject. NJEA President Wendell Steinhauer has expressed deep concerns about the implementation as well as the data of the new systems used in order

to inform decisions regarding teacher tenure and retention. “NJEA will vigorously represent any member who believes his or her evaluation is flawed or inaccurate,” he said. “This evaluation system is tremendously complex, and we will work to ensure that it is not misused to target or punish teachers unfairly” (Clark, 2015).

Given that the assessment systems are not content specific, questions have arisen about who makes the observation and resulting evaluation. In many cases, the responsibility for music teacher evaluation is the responsibility of a principal or other administrator who may or may not have any background in the arts. Additionally, the newly implemented evaluation models through which music teachers are assessed are potentially limiting in providing a complete consideration of the music educator’s abilities and contributions.

In order to further contribute to the literature regarding administrator and evaluation model suitability as independent areas of investigation for New Jersey, it was necessary to conduct a statewide study. The extent to which assessments of this type contribute to the professional growth of music educators in New Jersey had not been fully considered or understood prior to this research. In a political landscape where current tenure and evaluation reforms will affect music teacher retention, it was necessary to examine music teacher attitudes and perceptions of evaluation.

Up until this present study, administrator and evaluation model suitability as a part of music teacher evaluation have both been examined in a few states as well as one study in Canada (Becher, 2011 for Michigan; Geisler, 2014 for Ohio; Goddard, 2004 for Canada; Guerra, 2014 for Texas; Hirokawa, 2013 for Pennsylvania; Maranzano, 2002 for Virginia; Martin, 2014 in a combined study for Florida, North Carolina, Rhode Island,

and Tennessee; and Parulekar, 2014 for Ohio). Whereas previous studies have examined separately the variables of administrator and model type in the influence of music teacher attitude and perception of the evaluation process, a study had yet to be undertaken in which the interaction between contrasting administrator and model types was considered.

Two of New Jersey's most utilized and contrasting evaluation models at the time of this research were the Danielson Framework for Teaching (FfT) and the Stronge Teacher Effectiveness Performance Evaluation System (TEPES). Both are generic models and may seem to differ only in their construction and presentation, yet a fundamental difference between them is that the Danielson model evolved from what was believed to be effective teaching practice; Stronge came into being more directly as a tool for evaluation. This difference is key: the Danielson model on account of its instructional polarity is aligned closely to traditional classroom practices in which the observer is directed to the assessment of the verbal/cognitive mode; the Stronge model was conceived initially as an evaluative instrument and is less bound by any defined or explicit instructional practices or settings. Its non-polarized instructional orientation readily admits the consideration of instruction environments outside of the traditional classroom.

Evaluator type has already been recognized both as a concern and variable in music teacher evaluation research. In New Jersey as in other states, an administrator evaluating a music teacher may be a district arts supervisor with a degree in music, or may be a building principal or other administrator with little to no reference in the assessment of content specific considerations beyond classroom management and procedure.

Therefore, the presence of both contrasting administrator types and evaluation model types within New Jersey make it feasible to investigate an area yet to be undertaken within music teacher evaluation research: the interaction between the variables of evaluation model type with administrator type in influencing music teacher perceptions and attitudes of the evaluation process.

The intent of this study was ultimately to inform, and not to convince or advocate for changes within music teacher evaluation. Yet there may be perspectives that question the value of knowing music teachers' perceptions of the evaluation process in the first place, and how those perceptions might fit into an overall assessment of the evaluation process.

In response, this study revealed that 76 percent of the administrators who evaluated public school music teachers in New Jersey in 2016 (based on the sample obtained for this research) did not possess a music degree. This could be compared to a large corporation, factory, or other entity in which 76 percent of the supervisors of the workforce have no knowledge or experience in the areas they are evaluating, and the identical evaluation form is used in every instance to evaluate its workers regardless of their work or product. One continues this scenario with the consideration of the entity's annual expenditures, followed by the questions: How efficient is this entity, to what degree is it productive, how do its shareholders feel?

But returning to the question of how the perceptions of music teachers might fit into an overall assessment of music teacher evaluation, one could return to the above analogy and ask if there would be any value in considering the perceptions of the workers themselves? In examining the results of a recent Gallup Poll of workers' perceptions of

their supervisors, a Wall Street Journal article reported that “people don’t leave jobs, they leave managers” (Weber, 2015, para. 1). The study involving 7,200 adults found that roughly half had left a job voluntarily in order to get away from a manager. The absence of meaningful communication was cited as the chief reason. Parallels are seen in this study.

The Gallup Poll findings demonstrate the principle that if management and evaluation are to be improved, which ultimately represents a professional relationship whatever the sphere, then a valid approach is to understand the perspectives of those who are evaluated within it. The justification for this study therefore was to examine those perspectives within the sphere of music teacher evaluation.

Questions for Research

This research sought to answer three primary questions:

1. To what extent does administrator type (administrator with a degree in music or not) have a significant effect upon music teacher perception of the evaluation process in New Jersey?
2. To what extent does evaluation model type (Danielson Framework for Teaching and Stronge Teacher Effectiveness Performance Evaluation System) have a significant effect upon music teacher perception of the evaluation process in New Jersey?
3. To what extent does administrator and evaluation model type interaction have a significant effect upon music teacher perception of the evaluation process in New Jersey?

Limitations

The study was limited by examining music teacher perception exclusively from within the kindergarten to twelfth-grade public sphere within New Jersey. Findings were

not intended to translate to music teachers who taught within private, parochial, charter, or other settings, nor even intended to translate to other public school music teachers teaching in other states. Results, however, still may be applicable when or if similar evaluation practices are encountered in other environments, yet no conclusions were considered for environments that lay outside the parameters of this study.

In examining to what extent the Danielson and Stronge models in particular influenced teacher attitudes, a possible confound was that a district's particular culture may have influenced the attitudes and opinions that its teachers held towards evaluation in general and the model that they used in particular. Although the administrative course that a district takes is not normally within the domain of the music teacher it should be noted that such influence cannot be entirely ruled out. A provision was made to address this with two questions in the survey: 1) teachers were asked to what extent their district's evaluation model reflected their own evaluation philosophy and 2) if they were a member of the committee responsible for selecting the evaluation model used in their district.

It is not known to what degree or extent administrators (whether music or non-music) were trained in the intended and uniform use of evaluation models in their districts. This area was not controlled for in this research and represented a limitation. To claim validity in this area, it would have to be assumed that the representatives of a particular evaluation model or schoolwide personnel responsible for training administrators in the use of an evaluation model were consistent in the level and depth of training they provided, and that those who received the training used the model as originally intended.

Owing to the study's data collection, there was the slight possibility that the variable for administrator type might not reflect with 100 percent certainty some participant responses, yet the results obtained from this study for administrator type and suitability were consistent with previous music teacher evaluation research (Becher, 2011; Goddard, 2004; Guerra, 2014; Hirokawa, 2013; Maranzano, 2002; Martin, 2014; and Parulekar, 2014), and the study's sample size ($N = 308$) was sufficiently robust in admitting potential respondent anomalies.

Chapter 2

Review of Literature

Foundations of Teacher Evaluation

In general, teacher evaluation has been seen to fulfill at least four basic purposes: 1) teacher improvement, 2) school improvement, 3) personnel decisions, and 4) school status decisions (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). The first two involve improvement individually (teacher improvement) and collectively (school improvement). The second two involve accountability; individually (personnel decisions), and collectively (school status decisions). School improvement and school status decisions fall under the domain of program evaluation. Program evaluation has received specific consideration within music education (Ferguson, 2007). In addressing assessment and accountability, Richard Colwell writes, “Educational standards, content, performance, opportunity to learn, assessment, and teacher education are designed to provide a framework for accountability and to answer the question, Are the schools performing adequately for the dollars invested?” (Colwell, 2006, p. 204).

As to the nature of evaluation, distinctions have been drawn between evaluation and assessment. Colwell (2006) writes,

Evaluation is distinguished by the making of judgments based on data derived from measurements and other procedures, while assessment refers to a considerable body of data that has the potential to diagnose and provide clues to causes. Assessment is then used to improve or judge instruction or do both. (p. 206)

The Council of Chief State School Officers (CCSSO) only recognizes assessment and not evaluation in its influential *Model Core Teaching Standards and Learning Progressions for Teachers* (Council of Chief State School Officers, 2013, p. 48). Yet, prior to any description or consideration of assessment methodology, a reality is the operation of an external influence: decisions of policy in the determination of evaluative goals. Wise and his colleagues write, “the implementation of any school policy, including teacher evaluation policy represents a continuous interplay among diverse policy goals, established rules and procedures...all influence teacher evaluation procedures” (Wise et al., 1984, p. v). Linda Darling-Hammond suggests that the accountability, which assessment is intended to procure, has become political, legal, bureaucratic, professional, and market oriented (Darling-Hammond, Wise, & Pease, 1983). In consideration of market interests, the supply of 25 separate evaluation models currently approved for use in New Jersey alone (New Jersey Department of Education, 2015), and one particular model’s cost of “\$24 for every employee each year, plus a \$3,000 installation fee, and another \$3,000 a day for three days of administrative training” currently in use in New Jersey (Mooney, 2012, para. 31) supports such a justification.

Writers of supervision and assessment draw distinctions between at least two types of assessment: formative and summative. Formative assessment attempts to determine if an instructional goal is headed in the desired direction, and may offer suggestions if a correction of course is necessary, while a summative evaluation assesses if the goal or goals were ultimately met (Glickman, Gordon, & Ross-Gordon, 2010). Within teacher evaluation, formative assessments may be used to describe what is occurring in the classroom. It is conditioned on mutual agreement between teacher and

evaluator. Summative evaluation is a more externally imposed measure, uniformly applied, to judge all teachers on similar criteria. Both have ramifications for the music educator.

Two other types of assessment are recognized within instruction: interim and diagnostic. An interim assessment measures progress relative to a goal. It is a type of assessment that falls between formative and summative assessment that evaluates knowledge and skills relative to a specific set of goals. A diagnostic assessment is generally given prior to the beginning of a unit of study to evaluate prior knowledge (Education First, 2013).

In the observation of phenomena, whether they be events, behaviors, objects, or a music lesson, choices have to be made, whether willfully or not, as to how those events or things are to be considered. Writers on evaluation recognize at least two modes of inquiry, the quantitative and qualitative modes (Glickman et al., 2010).

Quantitative methodology. Asmus and Radocy (2006) define quantification as, “the assignment of a number to represent an amount or a perceived degree of something. That is, the association of numbers with behaviors, object or events” (p. 95). Within education, at least three distinct measures of data collection are available for the evaluator. They are 1) categorical frequency instruments, 2) performance indicator instruments, and 3) visual diagramming (Glickman et al., 2010). Each captures some quantitative measure in a manner that highlights a particular element to be considered. Statistics is frequently at the service of quantitative pursuits in order that relationships within and among collections of numerical data may be better understood.

Qualitative methodology. Qualitative considerations do not concern themselves in discovering meaning from discreet quantity as does the quantitative approach. A motto for the qualitative approach to observation could be: Leave your calculator at home. Qualitative observation, known also as descriptive evaluation, is likewise concerned with events and things, but they are not made to conform into a predefined category. Neither are they measured. In fact, Glickman describes a qualitative observation in the following: “The observer goes into the classroom with a general focus or no focus at all and records events as they occur” (Glickman et al, 2010, p. 246). Despite the surface appearance of a form of looking that may appear nebulous and ill-defined, qualitative inquiry is seen to be a rigorous, ordered, and disciplined mode of inquiry (Creswell, 2007). Returning to Glickman’s quote above, “the observer goes into the classroom...and records events as they occur,” it is the “how” in which events are recorded that determine the identity of a particular qualitative approach. Creswell (2007) for instance, identifies at least six approaches. Glickman (2010), for use within the educational realm specifically, identifies at least five. Each of the qualitative approaches identified by Glickman will be described very briefly:

- 1) The Verbatim and Selected Verbatim modes primarily record all verbal events and interactions taking place in the classroom. Afterwards, the observer identifies patterns or interpretations of behaviors.

- 2) In the Detached Open-ended Narrative mode, every person, event, or thing that attracts the attention of the observer is recorded. The evaluator begins with an empty page containing no categories, indicators or questions and records.

3) In the Participant Open-ended mode of observation, the observer becomes a functioning member of the classroom, assisting in instruction or participating as a student (Spradley, 1980).

4) A more specific form of qualitative observation is the Focused Questionnaire mode in which an observer seeks information about specific areas of inquiry, particular models of instruction, and learning goals (Marzano et al., 1988). The teacher evaluation models currently in use in New Jersey could be seen to broadly fall under this category.

5) Eisner (1985) developed a mode of inquiry known as Educational Criticism that merges detached and participant observation with description. Observers are trained to look at the classroom “as an art critic might look at a painting” (Glickman et al. 2010, p. 251). Evaluators through their familiarity with differing types of classrooms and forms of instruction become a sort of instruction critic, what Eisner refers to as “connoisseurship.”

When no observation system is available capable of collecting the desired data, tailored observation systems can combine both quantitative and qualitative modes.

A caution in the use of a qualitative approach as identified by Glickman (2010) is the need to be cognizant that one’s observation is affected by personal experience and values. This is even so with quantitative inquiry, if one acknowledges a tenet of quantum physics: that any event or phenomena cannot be measured without the measurement process itself interacting with the event and thus affecting the measurement (Greene, 2004).

The choice of a particular type of observation should ultimately depend on the focus of what is being observed (Creswell, 2007). According to Glickman (2010), it

should “put a mirror of the classroom up to the teacher, who can then attend to matters previously unknown” (p. 256).

The modes of inquiry—quantitative, qualitative, and tailored approaches available to evaluator and researcher alike offer the use of many lenses through which the world may be considered. The ultimate consideration, of course, should be not to lose sight of what one is observing. Yet as will be seen, policy as it influences teacher evaluation is often the ultimate determiner of which approach is used, and in some instances may not regard the appropriateness of a particular model in the consideration of a subject. What is seen to have primary importance is not what is seen, but how it is seen. The assessment/evaluation of music instruction is one example.

Music Teacher Evaluation

As a subject of focused quantitative and qualitative inquiry, the evaluation of the music teacher is a relatively recent development within music education research. Donald Taebel (1990a) claims that prior to his study examining the evaluation results of music teachers in Alabama, there was little known academic research on the evaluation of music teaching. It would take 12 years from Taebel’s initial research until music teacher evaluation would be considered as a dissertation topic (Maranzano, 2002). As for the relative lack of attention given to this field, Brophy (1993) suggests that “these types of studies require a considerable amount of work and time; they prove a daunting exercise for doctoral students who are interested in a more easily and quickly researched dissertation subject” (p. 14). The handful of dissertations that have appeared since Maranzano’s may support such a justification (Becher, 2011; Geisler, 2014; Goddard, 2004; Guerra, 2014; Hirokawa, 2013; Martin, 2014; Parulekar, 2014).

Another explanation could simply be: Who cares? Indeed, prior to 2009 the typical music teacher evaluation could afterwards find itself easily tucked away and soon forgotten in a permanent file. Yet this scenario would soon change. Between 2009 and 2014, more than 37 states had amended or changed their laws to make measures of student growth a significant part of teacher evaluation (Overland, 2014). Of the eight dissertations cited within this review, six were to appear in the years following 2009. Policy, it would appear, drives research as well as education (Elmore, 1992). On account of policy, Brophy rightly predicted as early as 1993 that the relative lack of attention given to the topic would eventually change.

Four Elements of Music Teacher Evaluation

If one examines the literature on music teacher evaluation from approximately 1990 to the present, one encounters themes that recur with varying degrees of frequency. Once found, themes may be identified and recorded with a note to reference. After an examination of the subject of music teacher evaluation, the following themes were identified. They are listed here alphabetically and not in order of appearance or frequency. They are 1) accountability, 2) administrator suitability, 3) effects of teaching too subtle to measure, 4) history of accountability measures, 5) instrument suitability, 6) policy, 7) process vs. result, 8) professional development, 9) professional concern, 10) self-evaluation, 11) standardized tests, 12) teacher compensation, and 13) value-added models. Table 1 provides examples of recorded themes and their sources as found in the literature on music teacher evaluation.

Table 1

Themes and Representative Sources

Author/Article	Year	Theme
Brophy, T. <i>Evaluation of music educators: Toward defining an appropriate instrument</i>	1993	Appropriate evaluation instruments Trained vs. untrained evaluators
Cody, A. <i>Alfie Kohn: We have to take back our schools</i>	2011	Reliance on corporate models, Value added models RTTT, loss of full independence within education The trump of global competitiveness directing the educational goals of a nation
Collins, I. H. <i>Assessment and evaluation in music teacher education</i>	1996	An agreed upon set of standards about what students should know and be able to do The entire arena of evaluation procedure Rater characteristics Appropriate evaluation instruments Frequency Location of evaluation (setting) Items in addition to instruments (portfolio) Impact of job security and salary Assessment impacting undergraduate music teacher preparation (lack of screening)
Colwell, R. <i>The status of arts assessment: Examples from music</i>	2003	Anxiety

Cope, P. <i>OFSTED, fun and learning: A case study of a school music inspection</i>	2003	Comparatively few pupils will enter conservatory or become professional musicians Appeal to standards vs. fun Rater Characteristics (one person evaluating—lack of objectivity; disciplined subjectivity) Accountability Snapshot evaluations
Doyle, K. O. <i>Evaluating teaching</i>	1983	Process vs. result
Duke, R. A. <i>Observation of applied music instruction: The perceptions of trained and untrained observers</i>	1987	Trained and untrained observers-no agreement among trained observers regarding quality and effectiveness of lesson; trained and untrained subjects vary considerably
Edgar, S. <i>Communication of expectations between principals and entry-year instrumental music teachers: Implications for music teacher assessment</i>	2012	Value added model Evaluator qualifications Mismatch between principal expectations and student performance or musical competency
Elmore, et al. <i>Curriculum Policy in Handbook of research on curriculum</i>	1992	Who decides curriculum
Elpus, K. <i>Merit pay and the music teacher</i>	2011	Teacher compensation
Fisher, R. <i>Debating assessment in music education</i>	2008	Music teacher accountability National Standards Need for national assessments Professional development as a result of evaluation Opposition to national assessment Need for national assessment

		Limited time to prepare for standardized tests
		Favoring performance over music content
		Criticism of concert-to-concert curriculum with no clear way to grade student progress
		Group assessment vs. individual assessment
		Standards created not for the benefit of music instruction but to satisfy policy makers
		Emphasizing narrow skills at the expense of content and understanding
Goddard, H. J. <i>Elementary music teachers' and principals' perceptions of music teacher evaluation</i>	2004	Administrative suitability
		Subject of evaluation relatively unexplored
		Capture of non-verbal behaviors in music education
		Evaluation of generic competencies vs. specific competencies
		Disagreement among principals and music educators as to who is best qualified to evaluate
		Generic forms, checklists, unidimensional qualities
		Peer coaching
		Mentoring
Guerra, A. <i>Perceptions of secondary level ensemble directors regarding teacher evaluations in Texas public schools</i>	2014	Purpose of evaluation
		A perceived inadequate administrator background

		Evaluation process of teachers actually benefitting student achievement
Henninger, J. <i>The effects of knowledge of instructional goals on observations of teaching and learning</i>	2002	<p>Administrator preparation</p> <p>Administrators unable to offer constructive suggestions to their teachers</p> <p>Music as a distinct and separate subject matter to be evaluated independently</p> <p>Administrator appropriateness; suggests that most administrators are reasonable confident in their ability to observe and evaluate music teachers appropriately</p>
Hourigan, R. M. <i>Race to the top: Implications for professional development in arts education</i>	2011	<p>Value Added Model</p> <p>Evaluator Suitability</p>
Iwanicki, E. F. <i>Focusing teacher evaluations on student learning</i>	2001	Most common models of evaluation appeared to do little to enhance teacher performance and failed to adequately recognize the unique contributions of teachers representing a wide variety of disciplines
Maranzano, C. <i>Evaluating music teachers in Virginia</i>	2002	<p>Need for evaluation in the music teacher community</p> <p>Firing of teachers when needed</p> <p>Managerial evaluation approach</p> <p>Limited applicability of evaluation instruments in common use</p> <p>Fixed models</p> <p>Self-evaluation</p> <p>Lack of qualifications of evaluators not trained in the field of music</p> <p>Better instruments, not administrators are the answer</p> <p>Existing broad theoretical studies appear to be inadequate for predicting the most</p>

salient or important music teaching practices

Absence of subject specific criteria

Lack of consensus over agreed-upon competencies

Evaluators with little training in music

Traditional parameter, “one-sized-fits-all” approach to evaluation not appropriate for music

Important results of music instruction are affective and are disconnected from traditional measurement

Consensus of content-little agreement among music educators regarding what should be taught

Process vs. product

Qualifications of evaluator

External evaluations of music programs, public performances, adjudicated festivals

Snapshot evaluation vs. feature film

NAfME Recommendations for Music Teacher Evaluation

2011

It is important for our field to closely monitor the area of policy work regarding music teacher evaluation

Developing of teacher evaluation instruments

Identification of three types of measures
1) student outcomes 2) teacher practices
3) combination of the two

NAfME Teacher evaluation position statement

2012

Basing music teacher evaluation on valid information based on student achievement that is directly attributable to the individual teacher

Evaluation instruments must be created to evaluate curriculum which is taught

Evaluation must be applied to the number of students taught and instructional time available

Assessment must work on multi-year cycle to allow for appropriate professional development and growth

Appropriateness of the instrument

Growth model

Evaluation conducted by trained musician

Orzolek, D. C. <i>The paradox of assessment: Assessment as paradox</i>	2006	Assessment of student learning; the term “assessment” is full of paradox
Overland, C. T. <i>Teacher evaluation and music education: Joining the national discussion</i>	2014	Value-added growth systems
		Paradigm shift from former mode of evaluation to new
		Teacher quality
		Direct evaluation no longer effective in dismissing ineffective teachers
		Inappropriateness of evaluation instruments
		Value added model – “instead of single snapshot, we will recognize progress and growth” -U.S. secretary of education Arne Duncan
		Teaching skills too subtle to measure quantitatively
Parulekar, M. S. <i>Determining criteria for the evaluation of high school band directors: A survey of high school principals and band directors in the state of Ohio</i>	2014	Adjudicated events as measures of effectiveness
		Need to define measures
		Evaluator qualifications

Perrine, W. M. <i>Music teacher assessment and Race to the Top: An initiative in Florida</i>	2013	History of accountability measures Are assessment instruments appropriate for performing arts Value added measures
Peterson, K. D. <i>Teacher evaluation: A comprehensive guide to new directions and practices</i>	2000	Process of evaluation in scholarly work rarely contributes to the professional growth and development of teachers Make evaluation a task managed by a teacher and not a thing done to a worker
Schmidt, C. P. <i>Individual differences in perception of applied music teaching feedback</i>	1989	Rater characteristics—personality variables of the evaluator effect evaluation perception and results
Schmidt, C. P. <i>Reliability of untrained observers' evaluation of applied music instruction</i>	1992	Inter-judge reliability Inter-rater reliability Evaluation of lesson based on verbal behavior only High agreement of evaluation in area of rapport Low agreement of evaluation in the area of discreet technical areas
Shuler, S. C. <i>Assessing teacher competence in the arts: Should Mr. Holland have gotten the gig?</i>	1996	Appropriate assessment instruments Supervision from non-arts administrators
Shuler, S. C. <i>Music education for life: Music assessment, part 2—Instructional improvement and teacher evaluation</i>	2012	Responsibility for professional growth Value added model Music policy at the bidding of policy mandates One-size-fits-all teacher evaluation model
Swanwick, K. <i>The “good-enough” music teacher</i>	2008	Teacher evaluation processes may be inappropriate Process vs. result

		Inappropriateness of measuring outcomes as a gauge to measure teacher effectiveness
		Inadequacy of instruments
Taebel, D. K. <i>An assessment of the classroom performance of music teachers</i>	1990a	Development of appropriate instruments
		Process vs. product
		Results of teaching are affective and cannot be measured
		Appropriate evaluation instruments in which teacher is not scored poorly for lack of verbal responses
		What training is necessary for an observer to have in order to accurately evaluate a music teacher
		How does one objectively evaluate the personal characteristics that research continues to reveal as essential for good music teaching
		Reliability of evaluation instruments
		Evaluating generic competencies
		Little agreement exists in how to define a "good teacher"
		Non-musician evaluators unable to evaluate effectively
		Evaluation systems that relies excessively on verbal exchanges and cognitive learning
		Value of cognitive and verbal exchange in evaluation
		Merit pay
		Teacher feedback in the effectiveness of

		the evaluation process, failure to include student performance as an indicator
		Prior to 1990 there is little known research on the evaluation of music teaching
Taebel, D. K. <i>Is evaluation fair to music educators?</i>	1990b	Is music teacher evaluation fair?
		Inadequate evaluation systems
Wise, A. E. <i>Teacher Evaluation</i>	1984	Teacher involved in the process of evaluation
		Schools must motivate and guide through the process of improving their own teaching
		Teacher evaluation as a means to improve instruction in the classroom for the improvement of student achievement

Upon consideration of how individual themes can relate one to the other and function within a governing framework, they can be seen to participate within one of three related groups or elements as they relate to the teacher. Together as a whole, these elements form and operate as a cognitive framework for music teacher evaluation. The elements of this cognitive framework are expressed here as *teacher*, *evaluation*, *evaluator*, and *result*. A justification for the choice of these elements as they relate one to the other is that they may be readily conceptualized into other spheres: in physics, for example, in the areas of matter, measure, action, and reaction; in art: subject, perspective, judgment, and worth; even in law: evidence, trial, verdict, and justice. In each, something is being considered by a measure with a resulting consequence. A survey of the literature of music teacher evaluation shows that its major considerations fall into one of four elements in which something (*teacher*) is being considered (*evaluator*) by a measure (*evaluation*) with a resulting consequence (*result*).

There are many common themes that appear in Table 1. These commonalities are reflected in Table 2. They are also grouped according to the elements identified above. It should be noted that the elements and themes contained in Table 2 resulted from an examination and consideration of the literature as a whole and not the other way around; care was taken not to examine the literature from a priori assumptions of the nature of music teacher evaluation. Rather, the literature created and confirmed the framework for music teacher evaluation found in Table 2. It will provide an outline for what follows in this chapter.

Table 2

Elements and related Themes

Teacher	Evaluator	Evaluation	Result
	Administrator suitability	Instrument suitability	Professional development
	Self-evaluation	Value-added models	Professional concern and perception
	Student/parent/peer surveys	Standardized tests	Teacher compensation

The use of a framework of music teacher evaluation is beneficial in the organization and consideration of the large amount of information on this subject. However, the use of a cognitive framework has the potential for not admitting considerations that may appear to lie outside of it (Bartlett, 1932; Piaget, 1928). Holding certain beliefs about music teacher evaluation also has the potential for interpreting situations that occur within it incorrectly as well. The very nature of music teaching itself,

how it is informed philosophically and influences the choices a music teacher makes is one such area. This has not been an area of interest in music education research thus far. Research has examined grade level and area of specialized subject area within music instruction as it influences evaluation (Edgar, 2012; Guerra, 2014; Maranzano, 2002; Taebel, 1990a). Yet, owing to the metaphysical nature of the philosophical assumptions that govern music instruction, a focused consideration of this topic as it relates to music teacher evaluation, although fundamental, is perhaps beyond the practical scope of this investigation. What shall commence rather, is a consideration of the elements of *evaluator*, *evaluation*, and *result* as they are encountered in the literature. These will ultimately relate to a culminating examination of evaluation practices in New Jersey. Together, they will inform questions for research.

Evaluator

Irma Collins (1996) delineated the considerations of music teacher evaluation into six areas of professional concern. They are 1) rater characteristics, 2) appropriate evaluation instruments, 3) frequency, 4) location of evaluation, 5) items in addition to instruments, and 6) impact of job security and salary as areas of professional concern (para. 28). Collins' first two areas, "rater characteristics" and "appropriate evaluation instruments" have since received the most attention in the literature since her formal identification of these areas in 1996 (para. 28). NAFME would ultimately draft its own list of areas in at least two position statements (MENC, 2010; NAFME, 2012). Indeed, the NAFME position statement may be seen as a distillation of the most current and prominent issues that appear within the literature of music teacher evaluation. NAFME's first two areas include 1) basing music teacher evaluation on valid information, and 2)

appropriateness of the evaluation instrument. Nine points later, NAFME places “must limit...evaluations to...individuals with adequate training in music” (p. 2) as last on their list. The literature overwhelmingly places this concern, administrator suitability, as first in frequency of appearance. This raises a question: is its remarkable position of last place of professional concern correctly interpreted as NAFME’s care to consider the potential volatility of teacher/administrator relations in delivering a challenge to administrative qualifications? Or in the end, is it simply the one point over which the music teacher will have the least control in addressing?

The concept of *Evaluator* is one of the elements in music teacher evaluation research. Its expression into language is encountered in a variety of forms within the literature and does much to indicate the nature and concern of the research. For example (listed chronologically) “trained and untrained observers” (Duke, 1987, p. 115); “differences in perception” (Schmidt, 1989, p. 110); evaluator “expertise” (p. 5), “competency” (p. 18), and “qualifications” (p. 19) (Taebel, 1990a); “trained and untrained evaluators” (Brophy, 1993, p. 7); “supervision from non-arts administrators” (Shuler, 1996, para. 29); “qualifications of evaluator” (Maranzano, 2000, p. 270); “incompetent and dubiously motivated individuals” (Cope, 2003, p. 313); “who should be involved in the process” (Goddard, 2004, p. 4); and “underprepared principals” (Edgar, 2012, p. 136) all appear with regular frequency within the literature. “Self-evaluation” (Shuler, 2012, para. 10); “self-peer reflections” (Orzolek, 2014, line 44), and “student/parent/peer surveys” (Doherty & Jacobs, 2013; Orzolek, 2014, line 80) have made recent appearances in the literature in part as a result of recent policy. What follows

will be a consideration of attendant themes of *Evaluator*, which include 1) administrator suitability, 2) self-evaluation/reflection, and 3) student/parent/peer surveys.

Administrator suitability. The history of *evaluator* studies within music teacher evaluation literature may be largely divided into two periods with studies appearing before 1990 examining rater characteristics and results within university research and those following examining *evaluator* in the arena of music teacher professional concern and efficacy.

Robert A. Duke's research (1987) on trained and untrained observers is a relatively early investigation into the nature of the evaluation of music instruction. Duke examined differences between trained and untrained observers' estimates of lesson duration devoted to student talk and performance, and teacher talk and performance. In comparing the results of music education and music therapy students with those of undergraduate education majors, Duke found that "both trained and untrained subjects vary considerably in their perceptions of observed events, even within highly specific and controlled circumstances" (Duke, 1987, p. 115). Additionally, Duke found that "observers may focus their attention to a much greater extent upon teacher behaviors as opposed to those of students" (p. 122).

Jacqueline Henninger (2002) investigating in a similar tenor of Duke's perception research also found that more experienced teachers, in contrast to novice ones, direct their focus of attention more towards student behaviors than on teacher behaviors. Henninger suggests that a major challenge for teacher educators is to "shift naïve observers' focus of attention away from trivial details, irrespective of their salience, and toward the critical variables that affect teaching and learning" (para. 8).

Another study is Charles P. Schmidt's investigation (1989) of evaluator personality variables as they affect evaluation perception and results. His study lends support to the hypothesis that the personality type of the evaluator can influence evaluation results. Schmidt's study examined extroverted and introverted personalities as they relate to the evaluation of music instruction in which different teacher feedback evaluative phrases were used. In the study, subjects viewed a music lesson in which various forms of prepared feedback were employed. Following the lessons, the subjects were administered the Myers-Briggs Type Indicator to determine their personality type. Schmidt's findings revealed that extroverted subjects when viewing and evaluating a lesson administered by an instructor exhibiting similar personality characteristics tended to rate the lesson more poorly than did subjects with more introverted personalities. Schmidt states, "While extroverted individuals may have a tendency to provide higher degrees of feedback as teachers, they may not value it when placed in the position of being the recipient of such feedback" (p. 199).

At the time Duke's research (1987) appeared on the differences between trained and untrained observers, the state of Alabama was preparing its own observers to assess the teaching of some 30,000 teachers, 500 of whom were music teachers. The observers for the Alabama Career Incentive Program, an incentive-based, merit-pay plan for teachers based on classroom performance were drawn from building principals and central office administrators. Following 35 hours of observer training, confirmation of rater agreement, and the development of instrumentation, evaluations were conducted on 10 competencies and a variety of classroom behaviors. The results showed that music teacher scores were below the mean score for all teachers for 7 of 10 competencies

(Wiersma, 1988). In light of these results, Donald K. Taebel posed that there could only be one of two explanations: either 1) “music teachers as a group are less competent than other teachers” or 2) the results indicate the model as defined by the competencies “is not serviceable for evaluating the performance of music teachers” (Taebel, 1990a, p. 17). Taebel concludes that instrumentation developed to assess teaching behaviors not normally observed or even appropriate for the music classroom and rehearsal was a source for the disparity of results. In Taebel’s assessment, “An evaluation system that relies excessively on verbal exchanges and cognitive learning may be inappropriate for music teachers” (1990a, p. 20). But it was in the questionnaire responses of the teachers themselves following evaluation, which uncovered the issue of validity of administrator competency. The music teachers were indeed more concerned with the competency of the evaluators rather than the instrument. The music teachers in the Alabama study completed the questionnaires prior to the evaluation results being published. Although only 33% of the music teachers expressed dissatisfaction over administrator qualifications, it is conjectured that the number would have been higher had the music teachers known how they had been rated as a population prior to completing the questionnaire. In related research on the evaluation of specialist teachers, Wise found in a survey of 32 school districts that nearly all of the respondents agreed that their principal lacked the sufficient competence to evaluate them accurately (Wise et al., 1984).

Taebel’s study of the Alabama Career Incentive Program is perhaps the earliest study documenting and analyzing observation results as it pertains to the professional concern of the music teacher. It is a “first shots fired” moment in the literature questioning the validity of music teacher evaluation as it pertains specifically to both

evaluation and *evaluator*. Whereas the studies of Duke (1987) and Schmidt (1989) investigated evaluation results, they were drawn in part from undergraduate and graduate music students as observers. Taebel's study, on the other hand, examined evaluation results conducted exclusively by non-music specialists within the professional arena. Its significance and influence in the literature is confirmed by subsequent reference in numerous articles and dissertations on music teacher evaluation following its appearance in 1990.

Following Taebel's study, mounting attention directed toward evaluator characteristics appears. Schmidt's (1992) oft-cited study "Reliability of Untrained Observers' Evaluations of Applied Music Instruction" considers specifically inter-rater and inter-judge reliability. Schmidt found high inter-rater reliability and agreement in areas surrounding teacher rapport. However, he found comparatively low agreement in more discreet technical areas such as "clarity of musical explanations" (p. 21), "flexibility" (p. 24), "teacher's identification of important musical problems" (p. 24), and "knowledge of instrumental and vocal technique" (p. 21). Although Schmidt states that "certain aspects of instruction by untrained observers might be relatively consistent between observers and stable within observers" (p. 18); areas of "rapport" such as "demonstration of patience, understanding" (p. 21) and "criticism vs. praise" (p. 22), he concludes areas of technical concern should not be considered for evaluations when they are conducted within limited time frames. Relying exclusively on the typical snapshot evaluation conducted by the non-music administrator suggests that receiving an accurate evaluation in content specific considerations is merely a matter of chance.

In the same year, Timothy Brophy (1993) published *Evaluation of Music Educators: Toward Defining an Appropriate Instrument*. Although he primarily addresses the concern of instrumentation following Taebel's earlier findings, he posits the question "What training is necessary for an observer to have in order to accurately evaluate a music teacher?" (p. 17). In a later publication, Collins (1996) will list "rater characteristics" ahead of "appropriate evaluation instruments" (para. 28) as her perceived main area of concern. Collins's position lends support to Taebel's earlier findings that music teachers in the Alabama study were more concerned with the competency of the evaluators rather than the instrument itself.

As early as 1996, Scott Shuler addressed administrator suitability even when the arts administrator possessed a music background. He acknowledges that many arts supervisors will lack content expertise in at least one of the arts areas in which they will evaluate teachers. To address the formation of the arts administrator who can provide informed assessment within all arts areas, Shuler mentions the existence of a few university programs that can provide such expertise to the arts administrator. Shuler states that more of these programs should be made available.

"A call for more inclusive models" is the subtitle of an article on music teacher performance evaluation put forth by Charles Maranzano in 2000. In it, he assesses the condition of the evaluative process with its use of "one-size-fits-all" (p. 268) approach, and the failure of the prevailing evaluative process to consider the "important results of music instruction which are affective" and are therefore "disconnected from traditional measurement" up to that time (p. 269). The issue of *evaluator* also appears in the article, yet is not brought into sharp relief.

Then in 2002, what may be considered as a signal event occurs within the literature: the appearance of Maranzano's *Evaluating Music Teachers in Virginia*. Its significance is not confined to its claim of being the first dissertation to treat public school music teacher evaluation (Maranzano, 2002, p. 51), but also that it is the first large-scale quantitative study to appear since Taebel's study (1990a) of Alabama public school music teachers 12 years earlier (Maranzano, personal communication, July 31, 2015).

An issue perhaps not of comparative importance in Maranzano's previous article (2000) is brought to the foreground in Maranzano's 2002 findings; similar to Taebel's findings in Alabama, evaluator concerns were the primary issue with Virginia music educators ahead of instrumentation. The majority of music teachers (45.9 percent) listed "evaluator expertise" as the "least favorable aspect of evaluation." Similarly, 48.9 percent listed "evaluator expertise" to the question "how can evaluation be improved," with only 3.6 percent listing "evaluator expertise" as their "most favorable aspect of evaluation" (p. 125).

Although Taebel reported only 33 percent of the Alabama music teachers expressing dissatisfaction with evaluator expertise 12 years earlier, there are distinctions between Taebel's and Maranzano's studies. Taebel's population was responding exclusively to the Alabama Career Incentive Program initiative in which evaluation results were not revealed to the teachers prior to their responses. Maranzano's population on the other hand, was responding to perceptions of their total evaluative experience within their careers up to the time they completed Maranzano's survey. Additionally, Taebel was examining responses from an entire population whereas Maranzano reported

survey results generated from a 69 percent return rate. At any rate, both studies raise two questions: 1) are Alabama and Virginia music teachers unique in their perceptions, or 2) are the views of Alabama and Virginia music teachers more reflective of the remainder of country? To arrive closer to an answer requires further investigation into the literature.

In a rebuttal featured in the *London Times*, Peter Cope (2003) criticizes Great Britain's Office for Standards in Education (OFSTED) for its perceived lack of objectivity in poorly rating a music teacher in an English school for putting too much "fun and enjoyment" (p. 307) into the curriculum at the perceived expense of standards. Although important considerations of what constitutes a music program is discussed in Cope's article, the nature of *Evaluator* and its relationship in the evaluation of a music program is viewed from multiple perspectives. Cope writes of the unfavorable OFSTED report that "It bases its critique on an appeal to standards, an appeal linguistically conceived to convey a sense of measurement and objectivity. These latter are absent—the report depends on the opinions of a single inspector" (p. 313). These words could resonate with at least 33 percent of Taebel's Alabama teachers (1990a) and the majority of Maranzano's Virginia teachers (2002). Furthermore, it is maintained that OFSTED reports in general represent "discipline subjectivity" (p. 313). According to Cope,

In practice, this amounts to an implicit claim to objectivity and there is no allowance made for the possibility of error or of legitimate professional difference...the grossly uneven power relations are all factors which imply the fragility and the potential injustice of imposed public judgments. (p. 313)

Goddard (2004) investigated evaluation perceptions of both teachers and administrators in a dissertation following Maranzano's study of Virginia music teachers.

Prior to Goddard's study, *evaluator* issues focused primarily on the views and perceptions of the teacher. Her research uncovered administrator perceptions as well. Disagreement between teachers and administrators as to whom were best qualified to evaluate music instruction was found to exist in the five Canadian school divisions examined by Goddard. She also found administrator focus on the non-technical areas of instruction at the expense of more content specific areas, and that music teachers provided a more comprehensive evaluation of music teaching than did the principals, whose behaviors have been documented in the prior research by Schmidt (1992). Her concluding recommendations based on her findings mirror Collins's own recommendations (1996, para. 28) published eight years earlier; both list evaluator issues as the number one area of concern, followed by instrumentation concerns, followed by concerns of frequency.

The mismatch between administrator expectations and the training music teachers receive is examined for the conductors of Michigan's secondary schools (Becher, 2011). According to Becher, many of the respondents felt that the evaluation process was not a good indicator of their abilities, and that administrators were not qualified to adequately assess or provide useful suggestions of their conducting, results which resonate with two statewide studies already mentioned (Taebel, 1990a; Maranzano, 2002).

Confounding the situation further is a perceived mismatch between evaluator expectations and beginning teacher assessment expectations. According to research conducted by Edgar (2012), "underprepared principals" focused primarily on "interpersonal and general teaching skills such as engaging students and classroom management, and not on musical expertise" when conducting evaluations (p. 136). Edgar

found that principals often did not adequately convey their expectations to entry-year instrumental teachers. He concludes that increased communication would benefit these teachers and could stem the high attrition rate in the profession.

In Edgar's research, Schmidt's earlier findings in 1992 once again come to the fore; there are definite parallels between Edgar's and Schmidt's conclusions and language. For instance Edgar's "underprepared" principals (p. 136) parallels Schmidt's "untrained observers" (p. 18); Edgar's findings of focus upon "interpersonal skills" (p. 136) parallels Schmidt's findings of high agreement in the area of "rapport" (p. 19). Edgar's findings suggest that administrators not possessing a music background may avoid considering discreet musical skills in favor of assessing general classroom skills (ones with which such administrators will be more familiar); Schmidt's findings (1992) suggest that when "untrained" (p. 19) observers are forced to consider discreet musical considerations with the use of a rubric, then there is low agreement among raters.

The early research of Duke and Schmidt also informs and gives relevance to Joy Hirokawa's (2013) comparison of evaluations by observers with varied levels of music background. She sought to examine the differences in the evaluations of music teachers when they are evaluated by individuals with varying backgrounds. Her research sought to answer the questions "to what extent do evaluation scores differ based on the type of observer" (p. 12), and "to what extent does a training session in the observation and evaluation of music teachers affect the actual evaluation completed by a participant with little or no musical background" (p. 13). Hirokawa's results showed no significant differences between evaluation scores as a result of training, but found "numerous subtle differences that suggested individuals with musical background were more attentive to

aspects specific to the music classroom” (p. v) such as content specific characteristics. Hirokawa found too, that most administrators perceive themselves as qualified in their ability to evaluate music teachers appropriately. Once again, administrators voice their confidence in their ability to evaluate music teachers appropriately (Hirokawa, 2013) contrasting with continued cries from music teachers who insist otherwise (Beaver, 2002; Guerra, 2014; Paurlekar, 2014).

The consideration of *evaluator* in the context of this review has so far been examined in two perspectives: 1) students as evaluators within university research, and 2) professionals as evaluators within school districts. Evaluator perspective appears in at least two more manifestations in music teacher evaluation literature as self-evaluation and surveys.

Self-evaluation. In self-evaluation, provision is made for teachers to reflect upon their own instruction and to evaluate themselves as a component within an evaluative framework (Danielson, 1996, 2012). Scott Shuler (2012) considers it “just as important for the independent growth of teachers as it is for students” (para. 10). A justification is that it allows teachers to take greater responsibility for their own professional growth. It is maintained that the evaluative process should ultimately be a task managed by a teacher, not a thing done to a worker (Peterson, 2000), and that teachers must be involved in the process of evaluation (Wise et al., 1984). Yet Shuler (2012) cautions that “in the absence of achievement data, it is far too easy to engage in self-delusion instead of self-assessment” (para. 12).

Teacher self-assessment/reflection is an established component within teacher evaluation. Danielson points to research demonstrating the benefits of “reflection” and

“self-assessment” (2012, para. 17) in providing the most effective vehicle for teacher growth. Self-assessment and reflection within music teacher evaluation have been documented and considered. In his study of Virginia music teachers, Maranzano (2002) reports that teacher self-assessment accounted for 39.9 percent of the process. In order to gain insight into music teacher instructional decisions, Diane Delaney (2007) sought to examine the content of music teachers’ evaluations of their own instruction as well as the instruction of other music teachers. Indeed, “self-evaluation” is a promoted component in a widely used evaluative model currently in use by school districts (Danielson, 2012). In view of the value-added models currently in use by those states that adopted Race to the Top (U.S. Department of Education, 2009), and the prominent weight given to standardized tests within that model however, self-evaluation’s influence on a music teacher’s evaluation in practice appears to be currently of minimal impact on teacher score. Doug Orzolek (2014) for instance reports that self/peer reflection accounts for a mere five percent of music teacher evaluation in a model developed in Minnesota. There appears to be minimal consideration of the topic within literature since 2009. As will be seen later, other pressing evaluative considerations in music teacher evaluation will far outweigh self-evaluation in importance since the implementation of Race to the Top mandates. In fact, NAFME’s 600-plus-word “Teacher Evaluation Position Statement” (2012) contains no provision or mention of “self-evaluation” as a necessary or desirable component in music teacher evaluation. While Danielson points to research demonstrating the benefits of reflection and self-assessment in providing the most effective vehicle for teacher growth (2012), music leaders are struggling that their own

teachers are even evaluated for the subject in which they teach (Shuler, 2012; Perrine, 2013; Overland, 2014).

Student/parent/peer surveys. The completion of student surveys is a regular part of teacher assessment at the collegiate level. New to the primary and secondary level as a result of Race to the Top guidelines is the inclusion of student, parent, and peer surveys in teacher evaluation (Doherty & Jacobs, 2013). States that adopt Race to the Top are given freedom whether to use the measure or not, as well as to what percentage it be allowed to affect a teacher's score. As of 2013, 17 states require or allow for a survey from a parent, student, or peer to be included in a teacher's evaluation (Doherty & Jacobs, 2013). Minnesota, who was slated to adopt student surveys by 2014 would allow 15 percent of a music teacher's evaluation to be determined by student surveys (Orzolek, 2014). The Bill and Melinda Gates Foundation funded *Measures of Effective Teaching* study (MET) found that although value-added analysis is more accurate in predicting success than any other single measure, student surveys included, MET researchers found a strong correlation between strong positive student learning outcomes and surveys in which a majority of students describe the learning environment as focused, engaging, and demanding (Bill & Melinda Gates Foundation, 2013). Of the 17 states that allow for parent, student, or peer surveys, only eight explicitly allow for the use of student surveys. It has been warned with specific reference to such surveys that any instrument is sensitive to wording and presentation, and that states and districts do not underestimate the resources required in the design of high quality and valid instruments (Doherty & Jacobs, 2013). The examination of such survey results is largely unexplored. Furthermore, research has yet to fully consider any relationship between student demographic data

and/or literacy scores with these specific types of survey results. The use of student surveys to rate music teachers is currently not a consideration in NAFME's teacher evaluation position statement.

Summary of *Evaluator*. Three primary positions emerge in consideration of *evaluator* in the literature: 1) music teachers who claim untrained administrators are not qualified to evaluate them, 2) untrained administrators who claim they are qualified to evaluate music teachers, and 3) research that claims that untrained administrators are both appropriate and inappropriate depending on what is being examined. Research supports the non-music background administrator in the reliable assessment of teacher/student rapport and classroom management, areas admittedly which could be encountered in nearly any classroom, irrespective of subject matter. The reliable assessment of content specific items pertaining to music by the non-music administrator has been found to be another matter.

In the attempt to arrive at a position closer to objectivity (regardless of evaluator experience), creators of evaluation models develop lenses, instruments through which predetermined competencies may be assessed. Such lenses are designed to allow predetermined teaching/learning behaviors to pass through, be recognized, and ultimately be measured. Teaching/learning behaviors not in consideration by the designer may not be captured or considered.

Evaluation

The concept of *evaluation* as it applies to instrumentation is an element in music teacher evaluation research. Similar to *evaluator*, its expression into language is encountered in a variety of forms within the literature and indicates the nature and

concern of the research. For example (listed chronologically) “reliable instruments” (Taebel, 1990a, p. 21); “capture of nonverbal behaviors” (Taebel, 1990a, p. 8); “generic competencies” (Taebel, 1990a, p. 6); “appropriate competencies” (Taebel, 1990b, para. 4); “inadequate systems, instruments” (Taebel, 1990b, para. 28) “appropriate instrument(s)” (Brophy, 1993, p. 1; Collins, 1996, para. 28); “one-size-fits-all” approach (Maranzano, 2000, p. 19; Shuler, 2012, para. 22); “snapshot evaluations” (Maranzano, 2000, p. 272; Cope, 2003, p. 314); “absence of subject specific criteria” (Maranzano, 2002, p. 125); “limited applicability of evaluation instruments” (Maranzano, 2002, p. 36); “music-specific criteria” (Goddard, 2004, p. iii); and “standardised itemised check-lists” (Swanwick, 2008, para. 10) appear with regular frequency within the literature. What follows will be a consideration of attendant themes of *evaluation* that include 1) instrument suitability, 2) value-added models, and 3) the use of standardized tests.

Instrument suitability. 1990 serves as a rough dividing line in the consideration of instrument concerns. Prior to 1990, instrumentation as reported in nearly all of the literature was developed to address specific research inquiries within university settings (Duke, 1987; Schmidt, 1989). In Taebel’s study (1990a) of the evaluation results of some 500 music teachers in Alabama however, instrumentation perception is investigated as it applies to active music teachers within the profession. In Taebel’s examination of evaluation criteria developed for the assessment of Alabama’s 30,000 teachers as a whole, he discovered an explicit instruction model as advocated by Madeline Hunter, Barak Rosenshine, and others in which 1) “orientation, presentation, guided practice, review, and assessment are the chief components of a lesson” and 2) “questioning is one of the most important teaching behaviors of a lesson” among other tenets (Taebel, 1990b,

para. 17) making implicit that the primary mode of teaching is the verbal cognitive mode. Taebel maintains that it was the result of the Alabama music teachers being assessed primarily through the lens of the verbal cognitive mode that resulted in the lower than average scores than for deficits of the quality of their instructions as it directly pertained to music instruction. In Taebel's words, "An evaluation system that relies excessively on verbal exchanges and cognitive learning may be inappropriate for music teachers" (p. 20). He further states,

Adequate evaluation of music teaching should be sensitive to both direct and indirect models of teaching, capture non-verbal behaviors by the teacher and students (including affect), account for sequencing, and measure teacher musicianship as well as the typical verbal behaviors of presenting questioning, and responding. (Taebel, 1990a, p. 8)

Taebel's findings are not limited to Alabama. Two years following the Alabama study, Taebel (1990b) gathered information from states considered leaders in the field of teacher assessment, including Georgia, Florida, and Texas. In those states, too, music teachers were considered below the norm for the majority of competencies. In Taebel's examination of the competencies, he found strong consideration given to questioning skills and being "heavily dependent" (para. 18) upon process-product research directed primarily to the teaching of reading and math at the elementary level. According to Taebel (1990a), "states have proclaimed that the competencies used in their statewide assessment programs broadly apply to all grade levels and to all subject areas...in other words, these competencies are generic" (p. 6). This is perhaps the earliest manifestation of the evaluative catchphrase "generic competencies" found in the literature.

An important consideration in the development of statewide evaluation systems is the assurance of high observer agreement. This gives rise to explicit descriptions of a competence in terms of observable teacher actions. Taebel explains that although music teachers do ask questions that require answers, music teachers typically will request a behavior that the student cannot express verbally—teachers typically ask their students to perform. In Taebel’s examination of the competencies, he found this request difficult to classify according to the instrumentation in use at that time. In fact, Alabama’s system was only one of the four that included the item “elicits performance” (Taebel, 1990a, p.14). In the Alabama results, music teachers were observed eliciting performance on 63 percent of the observation records in contrast to other teachers who were observed “eliciting performance” only 9 percent of the time. Taebel criticizes the committee that drafted the competencies on several points, which include the committee’s failure to specify the development of motor skills as a form of instruction in the case of music. Taebel concludes,

We believe that evaluation systems, in general, have been based on a model of teaching that stresses verbal behaviors by both teachers and students—a model that may be appropriate for math and reading but that does not include important dimensions of music teaching. In short, the model is not valid when used to evaluate the competencies of music teachers. (Taebel, 1990b, para. 20)

What follows in the literature is specific attention given to the area of instrumentation with a general recasting of Taebel’s concerns, conclusions, and recommendations.

Timothy Brophy’s *Evaluation of Music Educators: Toward Defining an Appropriate Instrument* (1992) is a forerunner of this type of follow-up literature.

Brophy's contribution (1993) appears in an educational landscape yet to be influenced by future federal mandates such as No Child Left Behind. He calls into question the appropriateness of the state created evaluative instruments, with their observer agreement influenced generic features emphasizing the verbal cognitive mode. In different language, he echoes much of Taebel. For example, "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?" (Brophy, 1993, p. 1) Though an extension of Taebel's work, the significance of Brophy's contribution lies in his synthesis of the literature up to that time towards the generation of "an appropriate instrument" (p. 1). He uses the words "need," "essential task," and "urgent" (p. 17) to describe the situation. Yet what may be seen as a disappointing result of the article is that Brophy in the end does not offer one. What he does offer however, are "broad areas which would serve as defining attributes" (p. 16) towards its creation. They include 1) "personal characteristics," 2) "musical competence and performance skill," 3) "effective use of nonverbal strategies," and 4) "effective planning for concept learning and aesthetic" (p. 16) among others; to paraphrase Taebel (1990a), "An evaluation system that [does not] rely excessively on verbal exchanges and cognitive learning" (p. 20).

The obstacle as identified by Brophy (1993) was the music teaching profession's lack of fundamental agreement on content—what should be taught. Even three years following Brophy's 1993 article, Irma Collins (1996) would be writing "no uniform national or state system of music teacher evaluation has been developed that answers

these questions, nor is there an evaluative instrument that a music teacher can study and practice to improve individual teaching” (para. 29). Collins, too, credits lack of agreement as the source of the problem. “Until we in the profession can agree on what we want students to learn about music (such as through the National Standards), the evaluation system will continue to be developed and administered by those outside the music education profession” (Collins, 1996, para. 29).

It has been observed that a challenge with any newly created educational reform is its successful and widespread implementation (Kertz-Welzel, 2008). It may be telling that Collins would still identify “lack of agreement” as an obstacle in the creation of instrumentation two years after the publication of the National Standards. MENC attempted to facilitate the process of implementation with the online publication of various guidelines and teacher materials. Another challenge the National Standards faced was its scope. The 1994 standards are comprehensive and delineated into nine separate content areas with the only the first three of the nine being performance based. Such a comprehensive approach to music education may not receive full implementation by secondary school music programs in which an emphasis may lie predominantly in performance (Kertz-Welzel, 2008). Elpus (2011) wrote,

the music National Standards failed to significantly alter the landscape of music education in the schools because they lacked enforcement of any kind (being dubbed “voluntary,” as were the standards released in all other subject areas) and because of the charged political rhetoric that surrounded the entire effort to create National Standards in all subject areas. (p. 188)

Emphasis on performance makes only a minority of the standards even possible (Kertz-Welzel, 2008). Compounding the problem further is an apparent incompatibility of MENC's vision and the potential views of the local school board. "What makes a good music teacher?" Doug Orzolek posed this question during a visit to a state conference of school board members. The answer: "No complaints or letters from parents or students, good concerts, good trips, a strong pep band for the games, trophies and awards, and good numbers" (Orzolek, 2006, para. 6). As to what criteria a music teacher should be evaluated, school board members have their opinions, too. These criteria are not instructional behaviors found anywhere in the National Standards. Nor will they be observed in a classroom.

Eight years after the publication of the National Standards, and six years after Collins' suggestion that they provide for a uniform national or state system of music teacher evaluation, Charles Maranzano (2002) in his "Evaluating music teachers in Virginia" reports that almost three-quarters of responding school divisions (72%) did not indicate that any of the state-wide uniform performance standards and evaluation criteria were incorporated in the documents that he received for analysis (p. 108). He reports an overall absence of subject specific criteria, limited applicability of evaluation instruments in common use, and lack of consensus over agreed-upon competencies as describing the condition of instrumentation in Virginia. Instead, he finds instrumentation reflecting much of the *status quo* found in previous administrator behavior research. Maranzano writes,

Many of the school divisions in this study reported criteria in their evaluation instruments that reflect teacher behaviors, such as classroom management

strategies and variety of teaching methods prominent in the process-product literature...they fall short of emerging national views of teaching and learning interactions in the classroom. (p. 108)

Aside from a minimal infusion of standards, such findings appear little changed from Taebel's own findings in Alabama ten years earlier.

Returning to Collins (2006), she lists the need for suitable evaluation instruments as number two in her list of concerns following administrator suitability as number one (para. 28). NAFME's own "Teacher Evaluation Position Statement" (2012) published six years later would also list instrumentation as the number two area of concern, but places Collins' number one area of concern of administrator suitability in last place as was earlier seen. Incidentally, NAFME's number one area of concern as identified in 2012 is the need for teachers to be evaluated "based on student achievement that is directly attributable to the individual teacher" (p. 1). This is in reaction to the consequences of Race to the Top (2009) and its resulting value-added models that have had their impact upon music teachers in recent years (Shuler, 2012). The impact of value-added models will be considered later in this review.

Heather Jean Goddard (2004) in her study of principals and elementary music teachers would reveal more focus on generic areas such as classroom management and student participation than upon music-specific criteria. Terminology such as "generic forms," "checklists," and "unidimensional qualities" appear in a text that could have been written fourteen years earlier. Her recommendations based upon her findings include "music specific content" as her number two area of concern following "music consultants or music teachers as evaluators with the principal" (p. iv) as number one.

Keith Swanwick (2008) is in opposition to “standardised itemised check-lists” (para. 10), even though they be music based. He opposes the prediction of desired outcomes, which long lists of standards create. He is not alone in his assessment of prescribed music standards. In the words of Jon Finney, “our interest is in qualities far beyond the attainment of task criteria, for completing a task is in itself irrelevant to what I am thinking of as richer learning” (Finney, 2006, p. 2).

Summary of *Evaluation*. The consideration of music teacher *evaluation* in the literature may be viewed as a study of continuous catch-up, in which music leaders and teachers are in the position of reaction, retooling, and defense. Since 1990, the literature has seen the predominant use of generic models within the profession. The years following *Race to the Top* (2009) with the implementation of the value-added model of teacher evaluation have twice removed the music educator in catch-up—in the development of appropriate and suitable evaluation instruments; most attention in the literature since *Race to the Top* has been diverted away from more focused instrumentation efforts towards ensuring that music teachers are assessed for the content which they teach, and even in some cases the students which they teach. Since then, instrumentation considerations alone within the context of music teacher evaluation appear to be a moot point.

Result

The concept of *result* is one of the elements in music teacher evaluation research. The intended as well as unintended results of evaluation will be considered in this section. *Result* as it is found in the literature may be delineated into at least three general

areas. They are 1) professional development, 2) professional concern and perception, and 3) teacher compensation.

Professional development. The generic label of “professional development” is used here to organize a sphere that encompasses multiple considerations. Besides the nearly universal use of the label “professional development” as a result of teacher evaluation, the term “feedback” is also encountered in music evaluation research (Doerksen, 1990; Goddard, 2004, p. 6; Guerra, 2014, p. vi; Maranzano, 2002).

Wise and his colleagues suggest that teacher evaluation could be used to improve instruction, the improvement of student achievement being the ultimate goal. They indicated that in order for teachers to buy-in to the act of improvement, they must believe that the feedback they receive is useful. They must experience the sense that “a given course of action is both worthwhile and possible” (Wise et al., 1984, p. v). They proposed this could be done in at least two ways: that teachers be involved in the process of evaluation, and that schools motivate and guide through the process of improving their own teaching. Although Wise does not delineate a specific subject in particular to which his words apply, they may be seen as appropriate to the music teaching profession, presupposing suitable administration is available. Scott Shuler (1996) writes that a particular shortcoming within music teacher professional development is the lack of appropriate personnel. As has already been seen, Shuler draws distinctions between arts supervisors who lack content expertise in at least one of the arts areas in which they evaluate. If full professional development is to be in reach for the music teacher, then when arts administrators exist, it is ideal that additional training be made available to them.

At the time of Shuler's article in 1996, Irma Collins also wrote about professional development. She does not include professional development as an area of concern along with her other identified areas of rater characteristics and instrumentation as seen earlier (Collins, 1996). Neither does she link "professional" with development. Rather, it is her perceived need to reform music teacher preparation at the undergraduate level before potential teachers enter into the profession. Collins identifies assessment and evaluation issues that raise fundamental problems for undergraduate music teacher preparation. Collins suggests remedies for the undergraduate music program to better prepare future music teachers. Sixteen years later, Shuler (2012) would write that it is impossible to squeeze everything a teacher should know into a bachelor's program, that the professional life of the great educator runs "24/7 in part because there is always more to learn" (para. 5). Indeed, Collin's observations were made in an era predating sweeping policy change, digital technology, and universal use of the internet. Yet, there is a perception within the literature that undergraduate music teacher preparation alone is insufficient in order to become a successful music teacher. Conway (2002) in evaluating the pre-service music teacher preparation program at a large Midwestern university through an examination of the perceptions of beginning teachers, their mentors, and their administrators concluded, "The most valuable aspects of the teacher education program cited by the graduates were parts of the teacher education program that we in music education really have the least control over" (para. 32). Shuler (2012) has asked veteran teachers and preservice undergraduates, "What percentage of what you need to know to be a successful teacher do you learn as an undergraduate?" (para. 3) Shuler relates that most place their response between 10 and 50 percent. He explains that the "latter, wildly

optimistic estimate” (para. 3) of 50 percent is usually given by teachers-in-training, and that there is a “widening disconnect between the skills that beginning teachers possess and the musical interest of the children in their schools” (Shuler, 2012, para. 4).

The evaluative process should ideally result in some positive change regardless of the discipline. Whether positive improvement occurs or not appears to be a contentious subject in the literature. Some find that evaluation improves teaching (Weber, 1987; Tuckman, 1995); others find that it does not (Iwanicki, 2001; Peterson, 2000). They claim that the process of evaluation in scholarly work rarely contributes to the professional growth and development of teachers. Within music teacher evaluation, Maranzano (2002) found a similar view in Virginia. He reports “If the negative responses of music teachers to the research questions employed in this study concerning professional growth are indicative of the entire teaching profession, then the evaluation process itself is in need of wholesale adjustments” (p. 129). In nearby Pennsylvania, Hirokawa (2013) in a vignette writes of music teacher evaluation as being “a waste of everyone’s time” (p. 4). In Georgia, supervision was described as “non-existent” and “distorted” (Beaver, 2002, p. i).

The literature on music teacher evaluation reveals some of the more unpleasant results of evaluation. Cope (2003) writes of the evaluative process as ideally being properly reflective—an open and collaborative exchange rather than attempting to “shame everyone into making things better” (p. 314). He documented a school music program in England in which government officials perceived the program as placing too much emphasis on “fun and enjoyment” (p. 307) rather than on standards. Cope observes that such considerations do raise important questions about music in schools and about

how questions of value and differences of philosophy are handled by evaluative procedures. His chief criticism of the evaluative procedure in this instance is the manner in which it sought to achieve resulting improvement. Cope writes,

Improvements are not likely in a system which delivers subjective and punitive ‘quality’ snapshots and ignores or undervalues the efforts and progress the school has made in advancing the cause of music in a way that seems laudable to parents and to readers of the report. (2003, p. 314)

Swanwick (2008) writes of Great Britain’s *Youth Music* movement reaching more than 1.4 million young participants. Swanwick in his examination of the program concludes that the program’s evaluation process enhances the work of the music leaders, some 6,000 in all. He reports “the opportunity for professional development facilitated by Youth Music made a very positive contribution to the quality of the musical environments in which these music leaders were active” (para. 41). Here, Swanwick draws a distinction between *music leaders* and *music teachers*. A *music leader* is a category of music educator whose activities take place beyond the scope of the typical music classroom; they may not be certified music teachers or even regard themselves as educators; they wish to communicate their way of making music to others. Swanwick observes that the music leaders may have been right to resist any professional development which did not have an appropriate or relevant musical focus, or that satisfied a generic itemized checklist. This, according to Swanwick “leaves unresolved the issue of reconciling two very different paradigms, the one represented by current practice in schools and government requirements, and the other by the less tangible but important concepts of richer learning and tacit knowledge” (para. 42).

In the case of Youth Music in the U.K., Swanwick claims that its professional development effectiveness is due to it being unhindered from current evaluative practice found in public schools. Some of the literature considered so far points to the effects of policy thought to be responsible in negatively impacting what should be the results of evaluation, such as appropriate professional development. The Youth Music study offers a glimpse of a scenario in which a nationwide music program is autonomous from governmental policy mandate. A distantly related scenario exists in the U.S. in which a state chooses to maintain its educational autonomy from the U.S. Department of Education's Race to the Top. Texas is one example. Do perceptions of music teachers differ in states that do not participate in Race to the Top from those that do?

Adrian Guerra (2014) in a dissertation exploring the perceptions of music ensemble directors in Texas secondary schools examined their views of the purposes and procedures of the Texas Professional Development and Appraisal System (PDAS). A finding was that the Texas music directors believed that the evaluation process is intended to provide feedback so that effective teaching may result; they believed that effective feedback would lead to improved student outcomes. However, the participants in Guerra's study indicated that the feedback they received from their evaluations was not useful towards improving their teaching. Furthermore, Guerra found that some participants believed that their evaluators lacked the necessary background in providing feedback thought to be useful in their work.

Despite Texas' independence from RTTT, the perceptions found there appear little different from those reported by Donald Taebel (1990a) in Alabama some 24 years earlier. They also appear little different from perceptions reported in other RTTT

participating states that have been recently examined (Geisler, 2014 for Ohio; Hirokawa, 2013 for Pennsylvania; Martin, 2014 in a combined study for Florida, North Carolina, Rhode Island, and Tennessee; and Parulekar, 2014 for Ohio).

The consideration of Texas may give place for pause. In view of research, it raises the question “to what extent is music teacher assessment really a state-to-state consideration?” Although Texas is free from RTTT mandates, the music teachers surveyed in Guerra’s (2014) research express similar views as those in states that have adopted RTTT. Another question is raised: it is maintained that music teachers will be more likely to seek employment in areas that are perceived to exhibit more stable suburban teaching environments than others as a measure of self-preservation (Perrine, 2013). To what extent is this true? Also, according to Cochran-Smith (2007) the new approaches to teacher accountability are becoming a significant issue in teacher preparation, and is perceived as a pitfall in teacher education. To what extent does this determine where new teachers will choose to teach? These last considerations direct attention to the area of professional concerns as a result of music teacher evaluation.

Professional concern and perception. Other results of evaluation are found in the literature that do not directly relate to professional development; some are self-explanatory. Examples include merit pay (Cody, 2011, para. 10; Elpus, 2011, p. 181; Taebel, 1990a, p. 8); “impact of job security” (Collins, 1996, para. 28); “dismissal of teachers” (Maranzano, 2002, p. 30); “evaluation by observation” no longer effective in dismissing ineffective teachers (Overland, 2014, line 39); and “narrower curricula” (Elpus, 2001, p. 182; Hodsoll, 1998, p. 97). Even “anxiety” (Fisher, 2008, para. 18), “act

of aggression” (Pieper, 2009, p. 26), and “getting the wrong people off of the bus” (Collins, 2001, p. 41) are encountered.

Some Modern philosophers distinguish between at least two forms of looking; the state in which one becomes aware of matter and form in the process of acquiring intellectual knowledge (Pieper, 2009). In one, the mind is passive and receptive; the attention is not strained, one simply receives—the state of contemplation. Opposite to this is the act of observation that implies that “we are beginning to count, to measure and to weigh up” (Pieper, 2009, p. 26), words that aptly describe the evaluative modus. German philosopher Ernst Jünger defines this particular form of looking, of observation as a tense activity, an act of aggression upon that which is being observed (Pieper, 2009, p. 26). Had Ernst Jünger been a music teacher (or any teacher) today, would he have viewed routine evaluation as an “act of aggression?” As teacher assessment becomes more influenced by free market principles (Elpus, 2011), it is perhaps not inappropriate here to mention Jim Collins. In his bestselling book on corporate management titled *Good to Great* (2001), he writes of the necessity for an organization not to initially ask, “where is the company going?” but rather “who is going?” resulting in “getting the right people on the bus, the wrong people off the bus, and the right people in the right seats” (p. 41). Kenneth Doyle (1983) writes, “lurking beneath these external pressures are internal stresses as well. Faculty wonder if, through all of this evaluation, their worst and deepest fears will finally be realized; they will be found out” (para. 6). A Texas music teacher expressed, “when I get a walk-through that has not been announced, it always makes me feel like they are hoping to walk in and find me not doing my job” (Guerra, 2014, p. 75). Evaluation bogeyman aside, words such as “anxiety” (Fisher, 2008, para. 18), “paranoid”

(Guerra, 2014, p. 75), and “tense and filled with distrust” (Guerra, 2014, p. 98) used directly to describe a result of music teacher evaluation cannot appear without at least some justification.

Maranzano (2002) cites the potential limitations of the evaluative process when a teacher can remain ensconced in a position due to the efforts and expense required of a school board to dismiss them legally when they are found to be ineffective. Twelve years later, the paradigm shift in teacher evaluation created by *Race to the Top*, according to Overland (2014), still raises questions. He writes,

There is a disconnect between our ability to identify poor teachers and our ability to retrain or remove them. It remains to be seen how the professional evaluation of music educators—and of those educators practicing many other nontraditional teaching styles, for that matter—will fit in this new paradigm. (line 47)

Teacher compensation. Merit pay is a part of the new paradigm. The practice of merit pay in reference to music teacher evaluation in the literature is found to predate *Race to the Top* inducements (2009) by at least 26 years. The publication *A Nation at Risk* (Gardner, 1983) thrust the recommendation of merit pay as one means of improving education. In the words of the report, “Salaries for the teaching profession should be increased and should be professionally competitive, market sensitive, and performance based...superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated (p. 30). Shortly thereafter, the Alabama Career Incentive Program passed in May of 1985 provided for an incentive-based, merit-pay plan for teachers based on what was thought to be best practices of teacher effectiveness (Wiersma, 1988). The results of this program were to culminate in Taebel’s (1990a)

findings of administrator and instrument suitability published a few years later. Race to the Top has since captured the primary focus of merit pay considerations.

The single-salary-schedule has been the traditional method of compensating teachers. It has been thought to result adversely in the so-called “widget effect,” in which a teacher is presumed to be effective in part on academic attainment and evaluation results. If the members of an entire teaching force are considered equally qualified, then their value is ultimately little more than that of a faceless, interchangeable part, hence “widget” (Weisberg et al., 2009). Danielson (2012) writes likewise of the typical evaluation as rating teachers at the highest level “on every item, with no guidance toward improvement” (para. 1). In contrast, merit pay allows for a portion of a teacher’s compensation to be tied directly to student/teachers tasks and outcomes.

Merit pay is a contentious subject. It is believed that utilizing free market principles in the educational sphere leads to perverse incentives and unintended consequences (Elpus, 2011, p. 182; Cody, 2011). The justification for this is formulated by research methodologist Donald T. Campbell stating that “the more any quantitative social indicator is used for social decision making, the more subject it will be to distort and corrupt the social processes it is intended to monitor” (Campbell, 1975, p. 35). Examples of perverse and unintended consequences include goal distortion, a condition in which teacher compensation tied directly to test scores ultimately narrows the curriculum. Merit pay is believed to lead teachers into “gaming the system” (Elpus, 2011, p. 182). Teacher cheating on student standardized tests has been already documented (Jacob & Levitt, 2003). It is also believed that it has the potential to “disrupt the

inherently collaborative nature of teaching in deleterious ways” on account of in-school teacher competition (Elpus, 2011, p. 182).

According to Elpus (2011), “music education scholarship has unfortunately been silent on the issue of teacher compensation” (p. 181). Irma Collins (1996) listed “impact of security and salary” (para. 28) as her sixth area of professional concern in music teacher evaluation during that time. Fifteen years later, NAFME (2011) makes no explicit mention of compensation in their teacher evaluation documents. Elpus cites the existence of chiefly theoretical considerations of merit pay in the literature rather than practical ones as the reason for its lack of focused consideration. Aside from a documented decrease in absenteeism of teachers eligible for merit pay incentives, Elpus suggests, based on research, that “merit pay does not significantly improve student performance” (2011, p. 184).

Summary of *Result*. The presence of professional development or feedback, teacher compensation, and teacher perceptions resulting from music teacher evaluation within the literature show that *result* is an element of music teacher evaluation research.

Although participating states are given leeway within Race to the Top guidelines in how their teachers are evaluated (with some states opting out entirely), a consistent perception from research emerges: music teacher evaluation does not result in professional development or feedback that is considered useful within states that research has examined so far. This was likewise found in a non-participant RTTT state. Music teacher evaluation judged to positively effect further professional growth was found to come out of a nationwide music program independent of government evaluative oversight.

The introduction of free market influenced inducements into the evaluative process is thought to pervert the very system it is intended to benefit. Adverse effects include narrowing of the curriculum, teachers gaming the system, and teacher insulation as a result of competition. Research in these areas has been more theoretical than practical.

Teacher Evaluation Models in New Jersey

Overview. New Jersey's \$38 million Race to Top award resulted in legislation known as TEACHNJ Act, which was signed into law in August of 2012. A component of this legislation centered on evaluation reform. A mandate included the creation or adoption of teacher evaluation systems that are based on teacher practice as well as multiple measures of student learning that differentiates between levels of performance as well as providing feedback for professional support and development (Shulman, 2012, p. 7). New Jersey's support system for this initiative is an online resource known as AchieveNJ. According to its website, its overarching mission is to better align educator evaluation with practices that lead to improved student outcomes (New Jersey Department of Education, 2015a).

Peter Shulman, Chief Talent Officer for the Division of Teacher and Leader Effectiveness within the New Jersey Department of Education identifies terminology as it refers to the implementation of New Jersey evaluation reform: a *teaching framework* refers to the philosophy, tools, and processes used to evaluate educators, while the *teaching practice observation instrument* itself is the specific instrument used to assess observable teacher or administrator practice (Shulman, 2012, p. 2). Adopted or district-created instrumentation must include 1) competencies—the specific indicators of

teaching practice that are assessed by a given teaching practice evaluation framework, and 2) evidence-supported teaching practice—scales or dimensions that capture multiple and varied aspects of teaching performance that must be attested by “knowledgeable practitioners or experts in the content prior to use in the observation of a teacher’s practice” (Shulman, 2012, p. 8). The instrument also must be validated through published research so that findings may be subject to professional peer review. Finally, the evaluation system must be shown to result in teacher growth and professional development (Shulman, 2012, p. 8). Instruments must include ratings in one of four category assignments: 1) ineffective, 2) partially effective, 3) effective, or 4) highly effective. A technological requirement for each model is that it be supported by a data management system—an electronic or internet-based platform for storing, organizing, analyzing, and reporting evaluation data. As of 2015, the New Jersey Department of Education has approved 25 such teaching frameworks for use within New Jersey Schools (New Jersey Department of Education, 2015b). They are listed below alphabetically in Table 3.

Table 3

Evaluation Models Approved for Use in New Jersey as of 2015

Charlotte Danielson: The Framework for Teaching
 Classroom Assessment Scoring System
 Classroom Strategies Scale Model
 Focal Point Teaching practice Model
 IMPACT
 H.E.A.T./Danielson Teacher Evaluation Instrument
 Insight Core Framework
 Kenilworth Teacher Evaluation Instrument
 Lenape Regional Teacher evaluation Instrument
 Marzano’s Causal Teacher Evaluation Model
 Mid-continent Research for Education and Learning Teacher Evaluation Standards (McREL)

North Star-Academy Teacher Evaluation Rubric
 Pearson Framework for the Observation of Effective Teaching
 Rhode Island Model: Teacher Evaluation & Support System
 Stronge Teacher and Leader Effectiveness Performance System
 Teacher Evaluation and Improvement Instrument
 The College-Ready Promise Teaching Framework
 The 5D+ Teacher Evaluation Rubric
 The Marshall Rubrics
 The Newark Public Schools
 The New Jersey LoTi Teacher Evaluation
 The SmartStart TeachElite Evaluation System
 The Thoughtful Classroom Teacher Effectiveness Framework

Of these 25, the five models most used in New Jersey are represented below in Table 4. This data was obtained from 469 school districts that participated in a recent survey (Mooney, 2013).

Table 4

Five Most Popular Teacher Evaluation Models in New Jersey as of February, 2013

Model	Number of districts reporting	Percentage
Charlotte Danielson Framework for Teachers	291	60%
Stronge+ Teacher and Leader Effectiveness System	53	11%
Mid-continent Research for Education and Learning Teacher Evaluation Standards (McREL)	45	9%
Marzano's Causal Teacher Evaluation Model	44	9%
The Marshall Rubrics	32	6.5%

As early as 1983, Darling-Hammond and her colleagues suggested that accountability and assessment in education are both market oriented. In the current teacher assessment climate, creators of evaluation models openly promote and vie for acceptance of their models at workshops and conventions. In an article titled "School Districts Comparison Shop for Teacher Evaluation Systems" (Mooney, 2012), James

Stronge, creator of his eponymous evaluation system told representatives from dozens of districts in New Jersey that his model was “the right tool for the task” (para. 1), but graciously added that his main competitors would not be a bad choice either, closing with the words, “I’m not at all biased...but this is the best tool and I hope you choose it” (para. 3). The competition between evaluation models and their creators for acceptance within the sphere of free enterprise gives place to the management maxim that one has to be either “number one or number two in each market, or exit” (Collins, 2001, p. 69).

Considered here will be the two most used evaluation models currently in use in New Jersey from a statistical standpoint: Danielson’s Framework for Teaching garnering a 60 percent acceptance rate, and Stronge’s Teacher and Leader Effectiveness System garnering an 11 percent acceptance rate (see Table 4). Each system is a conception of identified areas of teaching expressed through the creation of a cognitive framework comprised of chief components followed by smaller ones. These components broadly include preparative delivery of instruction, instruction and rapport, in addition to non-instructional professional responsibilities. The identifying features of the Danielson and Stronge frameworks appear at first glance to lie more in how each model is expressed rather than any great philosophical differences between them. At a presentation of evaluation models to district principals and administrators, a retired principal commented, “whether it’s this one or another like Danielson or Marzano...there is not one that goes, ‘Wow, there’s something new’...but it comes down to the presentation” (Mooney, 2012, para. 32).

Both Danielson and Stronge utilize large-scale components in the makeup of their overall frameworks. Danielson identifies her components as “domains” of which she

recognizes four: 1) planning and preparation, 2) classroom environment, 3) instruction, and 4) professional responsibilities (Danielson, 2014). Stronge recognizes seven components, what he identifies as “performance standards.” They are 1) professional knowledge, 2) instructional planning, 3) instructional delivery, 4) assessment of/for learning, 5) learning environment, 6) professionalism, and 7) student progress (Stronge, 2012a). Each model further delineates its major considerations into more discreet areas. Danielson divides her four domains into 22 separate components, and within each component there are one to five elements, making a total of 76 smaller elements. Whereas the Danielson model is built up of three descending layers of 1) domain, 2) component, and 3) element, the Stronge model is made up of two, with each major standard receiving four to nine performance indicators each. Therefore in their relative constructions, the Danielson model on the surface appears compact whereas the Stronge broader. The Danielson model is more vertical and the Stronge system is more horizontal. As to the models being subject to peer review, both the Danielson and Stronge systems are supported with their independent bodies of published research in support of their respective areas of reliability, validity, and effectiveness (Danielson, 1996; Xu, 2013).

Danielson Framework for Teaching. Danielson maintains that her four domains are aligned to the Interstate New Teacher Assessment and Support Consortium (InTASK) standards (Danielson, 2013a) of which there are 10: 1) the learner and learning, 2) content knowledge, 3) instructional practice, 4) professional responsibility, 5) application of content, 6) assessment, 7) planning for instruction, 8) instructional strategies, 9) professional learning and ethical practice, and 10) leadership and collaboration (Council of Chief State School Officers, 2011). Danielson’s Framework for Teaching (FfT) has

gone through four generations beginning in 1996, its first iteration intended to be a definition of what constitutes good teaching rather than a framework for evaluation.

Danielson writes,

I intended it to be a definition of good teaching, in all its complexity. I hoped (and wrote) that it might be useful for any number of purposes: first, and most importantly, for teachers' own self assessment and reflection; for teacher preparation, recruitment and hiring, mentoring and induction; for professional development; and yes, also teacher evaluation. The latter was simply one of many uses to which it could be put. (2013b, general questions about the framework)

According to Danielson, the 2007 edition sought to include frameworks for so called “non-classroom specialist positions” including school librarians, nurses, and counselors. Music as an instructional area was not considered. The 2011 edition was informed from the research findings of the Measures of Effective Teaching (MET) project, supported by the Bill and Melinda Gates Foundation. This resulted in the creation of tighter rubric language. The 2013 edition of FfT was in response to the instructional implications of the Common Core State Standards (Danielson, 2013). The language of the 2013 framework also appears more streamlined when compared to earlier editions. Table 5 displays Danielson's 2013 edition of Framework for Teaching (FfT). Her framework in this table appears as it does in her publications beginning in the top left-hand corner and read clockwise.

Table 5

Charlotte Danielson's Framework for Teaching, 2013 Edition

DOMAIN 1: Planning and Preparation	DOMAIN 2: The Classroom Environment
<p>1a Demonstrating Knowledge of Content and Pedagogy</p> <ul style="list-style-type: none"> • <i>Content and the structure of the discipline</i> • <i>Prerequisite relationships</i> • <i>Content-related pedagogy</i> <p>1b Demonstrating Knowledge of Students</p> <ul style="list-style-type: none"> • <i>Child and adolescent development</i> • <i>Learning process</i> • <i>Special needs</i> • <i>Students' skills, knowledge, and language proficiency</i> • <i>Students' interests and cultural heritage</i> <p>1c Setting Instructional Outcomes</p> <ul style="list-style-type: none"> • <i>Value, sequence, and alignment</i> • <i>Clarity</i> • <i>Balance</i> • <i>Suitability for diverse learners</i> <p>1d Demonstrating Knowledge of Resources</p> <ul style="list-style-type: none"> • <i>For classroom use</i> • <i>To extend content knowledge and pedagogy</i> • <i>Resources for students</i> <p>1e Designing Coherent Instruction</p> <ul style="list-style-type: none"> • <i>Learning activities</i> • <i>Instructional materials and resources</i> • <i>Instructional groups</i> • <i>Lesson and unit structure</i> <p>1f Designing Student Assessments</p> <ul style="list-style-type: none"> • <i>Congruence with instructional outcomes</i> • <i>Criteria and standards</i> • <i>Design of formative assessments</i> • <i>Use for planning</i> 	<p>2a Creating an Environment of Respect and Rapport</p> <ul style="list-style-type: none"> • <i>Teacher interaction with students, including both words and actions</i> • <i>Student interaction with students, including both words and actions</i> <p>2b Establishing a Culture for Learning</p> <ul style="list-style-type: none"> • <i>Importance of content and of learning</i> • <i>Expectations for learning and achievement</i> • <i>Student pride in work</i> <p>2c Managing Classroom Procedures</p> <ul style="list-style-type: none"> • <i>Instructional groups</i> • <i>Transitions</i> • <i>Materials and supplies</i> • <i>Performance of classroom routine</i> • <i>Supervision of volunteers and paraprofessionals</i> <p>2d Managing Student Behavior</p> <ul style="list-style-type: none"> • <i>Expectations</i> • <i>Monitoring student behavior</i> • <i>Response to student misbehavior</i> <p>2e Organizing Physical Space</p> <ul style="list-style-type: none"> • <i>Safety and accessibility</i> • <i>Arrangement of furniture and use of physical Resources</i>

DOMAIN 4: Professional Responsibilities	DOMAIN 3: Instruction
<p>4a Reflecting on Teaching</p> <ul style="list-style-type: none"> • <i>Accuracy</i> • <i>Use in future teaching</i> <p>4b Maintaining Accurate Records</p> <ul style="list-style-type: none"> • <i>Student completion of assignments</i> • <i>Student progress in learning</i> • <i>Non-instructional records</i> <p>4c Communicating with Families</p> <ul style="list-style-type: none"> • <i>Information about the instructional program</i> • <i>Information about individual students</i> • <i>Engagement of families in the instructional program</i> <p>4d Participating in a Professional Community</p> <ul style="list-style-type: none"> • <i>Relationships with colleagues</i> • <i>Participation in school and district projects</i> • <i>Involvement in culture of professional inquiry</i> • <i>Service to the school</i> <p>4e Growing and Developing Professionally</p> <ul style="list-style-type: none"> • <i>Enhancement of content knowledge and pedagogical skill</i> • <i>Receptivity to feedback from colleagues</i> • <i>Service to the profession</i> <p>4f Showing Professionalism</p> <ul style="list-style-type: none"> • <i>Integrity/ethical conduct</i> • <i>Service to students</i> • <i>Advocacy</i> • <i>Decision-making</i> • <i>Compliance with school and district regulations</i> 	<p>3a Communicating With Students</p> <ul style="list-style-type: none"> • Expectations for learning • Directions for activities • Explanations of content • Use of oral and written language <p>3b Using Questioning and Discussion Techniques</p> <ul style="list-style-type: none"> • Quality of questions/prompts • Discussion techniques • Student participation <p>3c Engaging Students in Learning</p> <ul style="list-style-type: none"> • Activities and assignments • Grouping of students • Instructional materials and resources • Structure and pacing <p>3d Using Assessment in Instruction</p> <ul style="list-style-type: none"> • Assessment criteria • Monitoring of student learning • Feedback to students • Student self-assessment and monitoring of progress <p>3e Demonstrating Flexibility and Responsiveness</p> <ul style="list-style-type: none"> • Lesson adjustment • Response to students • Persistence

In examining Danielson’s framework in relation to the evaluation of the music educator, elements in Domain 1 make presumptions on the nature of the administrator evaluating music instruction. Appropriate assessment of elements contained in 1a “Demonstrating knowledge of content and pedagogy” that include 1) knowledge of content and structure of the discipline, 2) knowledge of prerequisite relationships, and 3) knowledge of content-related pedagogy, would assume that the evaluator has a background in music instruction. Similarly, the element “knowledge of child and adolescent development” presumes that the evaluator would know and recognize if a pedagogical approach in music was appropriate for a particular child. These areas as they

specifically pertain to administrator suitability have been examined in music teacher evaluation commentary and research (Becher, 2011; Maranzano, 2002; Shuler, 1996; Taebel, 1990a).

In the 2007 edition, Domain 2's "teacher interaction with students" and "students' interaction with others" is potentially ambiguous as it is applied to music teacher assessment; it calls to mind the investigations of music teaching being sensitive to both direct and indirect models of teaching (Brophy 1993, Taebel, 1990a). In the 2013 edition, the added language, "including both words and actions" appended to teacher and student interaction directs the observer to potential kinesthetic activity within instruction, bringing that element closer to a performance consideration within music instruction.

Instructional areas within FfT mirror the generic type of considerations found in research discussed in this chapter. This is most evident in the instructional area of Domain 3 with component 3b listed as "using questioning and discussion techniques" with 1) quality of questions, 2) discussion techniques, and 3) student participation appended as elements. These reflect the verbal cognitive mode of instruction not normally found in the music classroom. "Use of oral and written language," "instructional materials and resources," and "grouping of students" encountered in this domain are also more grounded in traditional classroom practice than in the music classroom or rehearsal. An exception could be made in the case of instruction in the non-performance spheres of music curriculum, subjects such as music history, appreciation, or theory. The cognitive verbal mode of instruction as it relates to music teacher evaluation has been considered in the findings of Taebel (1990a) and others after him, all of which have been discussed previously in this chapter.

Assessment of the music teacher through the intended design of Domain 4, Professional Responsibilities, potentially impacts the music teacher, particularly in 4b “maintaining accurate records” including the elements of 1) student completion of assignments, 2) student progress in learning, and 3) non-instructional records. The issue of student performance records for every student in performing ensembles as they pertain to a teacher’s evaluation has been a subject of concern within music administrator circles (Perrine, 2013, Winerip, 2012).

As previously noted, Danielson has developed frameworks for specialty areas for the enhancement of professional practice and evaluation of school librarians and counselors. Teaching in music, theater, dance, art, and physical education have yet to be specifically addressed in FfT. As FfT specifically relates to specialty subjects, Danielson writes:

There’s no doubt that the *Framework for Teaching* must be considered in light of the “context” of the classrooms in which teachers are being observed, and that “knowing one’s students” is different, in practice, when a teacher teaches hundreds (as in music, PE, or art) from what it might be in a primary classroom with, say, 23 students. As in other aspects of using the FfT, it’s important for common sense and reason to prevail. Therefore, a vocal music teacher might know that the alto section is coming in too early at a specific point in a piece of music. That same teacher might also know, however, that a particular student has a strong voice that might be suitable for a small solo role. But much of the teacher’s knowledge of students will be, inevitably, group-based. (Danielson, 2013b, general questions about the framework)

Stronge Teacher Effectiveness Performance Evaluation System. The Stronge Teacher Effectiveness Performance Evaluation System (TEPES) is a component of a larger system known as the Stronge Teacher and Leader Effectiveness System. It contains components for hiring, developing, and supporting educators through a set of uniform, aligned criteria in addition to providing an independent system for teacher evaluation. The latest manifestation of TEPES is known as Stronge+, an enhanced version of the original model, the result of feedback obtained throughout the United States and abroad (Stronge & Tonneson, 2015). Regarding the genesis of the Stronge model, its creator writes,

A number of years ago I advocated a set of 20 or so standards for use in teacher evaluation. However, after field-testing this system with many organizations over numerous years, I became convinced that this design simply isn't practical. Thus, I moved to a more simplified set of standards that retain the diagnostic profile of teacher performance, with seven teaching performance standards included in the version of my system that is being piloted in New Jersey. (Stronge, 2012a, p. 3)

According to Stronge, a hallmark of his model is its flexibility; certain aspects of the model may be modified to meet specific school district needs. Additionally, the model does not use its performance indicators in a checklist-like manner as does the Danielson model with its 22 components and 76 elements. In the Stronge model, teachers are rated only on the seven performance standards. Virginia Tonneson explains:

The performance indicators are research-based examples of what it might look like if the teacher was successfully performing the standard. In other words, they are observable tangible "look-fors." However, in the Stronge model, we do not

use the indicators as a checklist. They are merely examples...the list is not intended to be all inclusive, and you would not expect every teacher to demonstrate every performance indicator. In fact, districts are able to modify the indicators to meet their specific needs or areas of focus. (V. Tonneson, personal communication, September 1, 2015)

Indeed, a selling point that Stronge makes for his model is that it is comparatively simplified when compared to other systems (Stronge, 2012b, p. 3). Another difference between the Danielson and Stronge models is that one was informed and evolved from what was believed to be effective teaching practice, the other coming into being more directly as a tool for evaluation. This difference is key: the Danielson model on account of its instructional polarity is aligned closely to traditional classroom practices; the Stronge model was conceived initially as an evaluative instrument and on the surface appears less bound by any defined or explicit instructional practices or settings.

Stronge's major areas are likewise aligned to InTASC (Stronge, 2012a). Stronge mentions explicitly of his model moving to a more simplified set of standards after being convinced that his initial design was not practical (2012a). A comparatively simpler design from the Danielson model just considered is evident on first glance in his two-tiered construction containing 7 performance standards followed by a total of 53 performance indicators. There initially appears to be less language in Stronge's framework when compared to the Danielson model; Danielson's effective teaching orientation resulted in more explicit references to classroom practices. Stronge, however, is less explicit in description of what should be observed. As a result, his model, although generic, makes itself perhaps more adaptable to the consideration of instructional

activities and practices which may occur outside of the traditional classroom. Whereas Danielson refers directly to verbal cognitive mode instruction, Stronge is more ambiguous. Table 6 displays Stronge's Teacher Effectiveness Performance Evaluation System.

Table 6

James Stronge's Teacher Effectiveness Performance Evaluation System

Performance Standard 1: Professional Knowledge

The teacher demonstrates an understanding of the curriculum, subject content, and the developmental needs of students by providing relevant learning experiences.

Sample Performing Indicators (Examples may include, but are not limited to the following.)

The teacher:

-
- 1.1 Addresses relevant curriculum standards.
 - 1.2 Integrates key content elements and facilitates students' use of higher-level thinking skills in instruction.
 - 1.3 Demonstrates an ability to link present content with past and future learning experiences, other subject areas, and real-world experiences and applications.
 - 1.4 Demonstrates an accurate, current, and deep knowledge of the subject matter and a working knowledge of relevant technology.
 - 1.5 Exhibits pedagogical skills relevant to the subject area(s) taught and best practices based on current research.
 - 1.6 Bases instruction on goals that reflect high expectations for all students commensurate with their developmental levels.
 - 1.7 Demonstrates an understanding of the intellectual, social, emotional, and physical development of the age group.
 - 1.8 Demonstrates an understanding of appropriate accommodations for diverse learners (e.g., English learners, gifted learners, students with special needs, etc.).
 - 1.9 Uses precise language, correct vocabulary and grammar, and acceptable forms of communication as it relates to a specific discipline and/or grade level.

Performance Standard 2: Instructional Planning

The teacher plans using the state's standards, the school's curriculum, data, and engaging and appropriate strategies and resources to meet the needs of all students.

Sample Performing Indicators (Examples may include, but are not limited to the following.)

The teacher:

-
- 2.1 Analyzes and uses multiple sources of student learning data to guide planning.
 - 2.2 Plans accordingly for pacing, sequencing, content mastery, transitions, and application of knowledge.
 - 2.3 Consistently plans for differentiated instruction.

- 2.4 Aligns lesson objectives to the school's curriculum and student learning needs.
- 2.5 Develops appropriate course, unit, and daily plans, and adapts plans when needed.
- 2.6 Plans and works collaboratively with others to enhance teaching and learning.

Performance Standard 3: Instructional Delivery

The teacher uses a variety of research-based instructional strategies relevant to the content area to engage students in active learning, to promote key skills, and to meet individual learning needs.

Sample Performing Indicators (Examples may include, but are not limited to the following.)

The teacher:

-
- 3.1 Builds upon students' existing knowledge and skills.
 - 3.2 Differentiates the instructional content, process, product, and learning environment to meet individual developmental needs.
 - 3.3 Motivates students for learning, reinforces learning goals consistently throughout the lesson, and provides appropriate closure.
 - 3.4 Develops higher-order thinking through questioning and problem-solving activities.
 - 3.5 Provides remediation, enrichment, and acceleration to further student understanding of material and learning.
 - 3.6 Communicates and presents material clearly, and checks for understanding.
 - 3.7 Communicates and presents material clearly, and checks for understanding.
 - 3.8 Communicates clearly, checks for understanding using multiple levels of questioning, and adjusts instruction accordingly.

Performance Standard 4: Assessment of/for Learning

The teacher systematically gathers, analyzes, and uses all relevant data to measure student progress, guide instructional content and delivery methods, and provide timely feedback to students, parents, and stakeholders.

Sample Performing Indicators (Examples may include, but are not limited to the following.)

The teacher:

-
- 4.1 Uses pre-assessment data to develop expectations for students, to differentiate instruction, and to document learning.
 - 4.2 Uses pre-assessment data to develop expectations for students, to differentiate instruction, and to document learning.
 - 4.3 Uses a variety of formal and informal assessment strategies and instruments that are valid and appropriate for the content and for the student population.
 - 4.4 Uses high quality questioning to gauge student understanding.
 - 4.5 Uses assessment tools for both formative and summative purposes to inform, guide, and adjust students' learning.
 - 4.6 Collaborates with others to develop common assessments, when appropriate.
 - 4.7 Aligns student assessment with approved curriculum standards and benchmarks.
 - 4.8 Collects and maintains a record of sufficient assessment data to support accurate reporting of student progress.
 - 4.9 Communicates constructive and frequent feedback on student learning to students,

parents, and other stakeholders (e.g. other teachers, administration, community members, as appropriate).

Performance Standard 5: Learning Environment

The teacher uses resources, routines, and procedures to provide a respectful, positive, safe, student-centered environment that is conducive to learning.

Sample Performing Indicators (Examples may include, but are not limited to the following.)

The teacher:

-
- 5.1 Arranges the classroom to maximize learning while providing a safe environment.
 - 5.2 Establishes clear expectations, with student input, for classroom rules and procedures early in the school year, and enforces them consistently and fairly.
 - 5.3 Maximizes instructional time and minimizes disruptions.
 - 5.4 Establishes a climate of trust and teamwork by being fair, caring, respectful, and enthusiastic.
 - 5.5 Encourages student engagement, inquiry, and intellectual risk-taking.
 - 5.6 Promotes respectful interactions and an understanding of students' diversity, including language, culture, race, gender, and special needs.
 - 5.7 Actively listens and makes accommodations for all student needs, both intellectually and affectively.
 - 5.8 Promotes an environment that is academically appropriate, stimulating, and challenging.

Performance Standard 6: Professionalism

The teacher maintains a commitment to professional ethics, communicates appropriately, and takes responsibility for personal professional growth that results in the enhancement of student learning.

Sample Performing Indicators (Examples may include, but are not limited to the following.)

The teacher:

-
- 6.1 Adheres to federal and state laws, school policies, ethical guidelines, and procedural requirements.
 - 6.2 Maintains positive professional behavior (e.g., appearance, demeanor, punctuality, and attendance).
 - 6.3 Incorporates learning from professional growth opportunities into instructional practice and reflects upon the effectiveness of implemented strategies.
 - 6.4 Identifies and evaluates personal strengths and weaknesses, and sets goals for improvement of personal knowledge and skills.
 - 6.5 Engages in activities outside the classroom intended for school and student enhancement.
 - 6.6 Works in a collegial and collaborative manner with administrators, other school personnel, and the community to promote students' well-being and success.
 - 6.7 Builds positive and professional relationships with parents through frequent and appropriate communication concerning students' progress.
 - 6.8 Serves as a contributing member of the school's professional learning community through collaboration with teaching colleagues.

- 6.9 Uses precise language, correct vocabulary and grammar, and acceptable forms of oral and written communication.

Performance Standard 7: Student Progress

The work of the teacher results in acceptable, measurable, and appropriate student academic progress.

Sample Performing Indicators (*Examples may include, but are not limited to the following.*)

The teacher:

-
- | | |
|-----|--|
| 7.1 | Sets acceptable, measurable and appropriate achievement goals for student progress based on baseline data. |
| 7.2 | Documents the progress of each student throughout the year. |
| 7.3 | Provides evidence that achievement goals have been met, including the state-provided growth measure when available as well as other multiple measures of student growth. |
| 7.4 | Uses available performance outcome data to continually document and communicate student progress and develop interim learning targets. |
-

As noted earlier, Stronge’s framework does not direct the evaluator to focus upon specific indicators. In the first two areas of “Professional Knowledge” and “Instructional Planning” for example, it directs the evaluator towards appropriate curriculum and state standards rather than detailing specific teaching practices that may potentially limit the model’s use in instructional areas outside of the traditional classroom. Similar to Danielson, being a generic instrument the Stronge model makes assumptions that the observer is capable of recognizing that a music teacher “demonstrates an accurate knowledge of the subject matter” (1.4). Yet, in Performance Standard 3: Instructional Delivery, there is nothing explicit that directs the evaluator towards classroom pedagogical practice. In the Performance Standard 4, Assessment of/for Learning, the evaluator again is directed to an existing curriculum and accompanying state or other standards in the confirmation that an assessment is properly aligned to subject matter. This could be potentially challenging when used in a district that does not employ a content supervisor. In such cases, a body of teachers within a discipline may be

comparatively more autonomous in its curricular decisions and direction. The practices of music teachers without district supervision have been examined in this regard: Pierce (2005) in a study examining music teachers in northern New Jersey who teach independently, “without some entity coordinating a music program and holding the teachers accountable to the established curriculum, teachers will tend to act in isolation with little regard for their places in the curricular continuum” (p. 11).

Professional standards 5 and 6, Learning Environment and Professionalism likewise are not aligned to any detailed classroom practices. Stronge’s model could be described as a “more generic” of the generic models in that its wording is careful to eschew any specific instructional/learning environments. Whereas Danielson makes specific reference to interaction between teacher and student, as well as student to student interactions via prescribed elements within the environmental domain (Domain 2a), Stronge appears intentionally conscious of not providing any specific reference to practice and space, thereby creating a lens which does not direct the observer to identify specific and thereby potentially limited environmental or instructional considerations.

Stronge provides a separate area for student progress (Performance Standard 7) in which the work of the teacher results in what he describes as acceptable, measurable, and appropriate student academic progress (Stronge, 2012a). He specifically refers to providing evidence that achievement goals have been met, including state-provided growth measure when available as well as other multiple measures of student growth. This provision clearly addresses current RTTT policy in mandating the use of value-added models and student growth objectives in the creation of student assessments, the results of which figure into the assessment rating of a teacher.

Other Teaching Frameworks. Two other models are used in at least nine percent of New Jersey school districts respectively. They are Mid-continent Research for Education and Learning Teacher Evaluation Standards (McREL) and Marzano’s Causal Teacher Evaluation Model (see Table 4). Both include competencies and evidence-supported teaching practices as prescribed by TEACHNJ legislation, their differences lying partly in their respective presentations.

The McREL large components are called standards of which there are five: 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 facilitated learning for their students, and 5) Teachers reflect on their practices. Already, an identifying difference between McREL and the two models discussed previously is evident: Danielson and Stronge’s first major component targets preparation of instruction whereas McREL’s targets an area identified specifically as leadership. Leadership within the McREL model is a focused consideration, connected to the idea that teachers should take responsibility for the future of their students and that students are “prepared for life in the 21st century” (Williams, 2009, p. 2). Danielson and Stronge do not neglect leadership, yet place their leadership considerations towards the end of their models expressed as smaller elements. Contrasting with the prominence that Danielson and Stronge give to planning and preparation (placing this area in their first domains), McREL addresses this area further down in its third standard. Similarly to Danielson and Stronge, McREL delineates its major areas into smaller ones, what McREL labels as elements of which there are 25 in all. Within the McREL system, the evaluator is drawn to consider these smaller elements in a checklist-like fashion during the evaluation

process, a modus similar to Danielson with its 22 components, yet dissimilar to Stronge in which its performance indicators are not intended to be all inclusive.

Marzano's Causal Teacher Evaluation Model is comprised of 4 domains followed by 60 elements contained within areas called segments. The domains are 1) Classroom Strategies and Behaviors, 2) Planning and Preparing, 3) Reflection on Teaching, and 4) Collegiality and Professionalism (Marzano, 2013). Similar to Danielson, the Marzano model was informed and evolved from what was believed to be effective teaching practice. Marzano considers his model, "an aggregation of the research on those elements that have traditionally been shown to correlate with student academic achievement" (Marzano, 2013, p. 1).

A feature of the Marzano model is the identification of "on the spot" behaviors that clearly emphasize what the model states should occur in the classroom. They put into sharp relief discreet teaching behaviors aligned to classroom practice. Examples of "on the spot" behaviors include "using academic games," "using physical movement," "providing opportunities for students to talk about themselves," and "asking question of low-expectancy students," actions that may not normally be observed in the music rehearsal. Another identifying feature of the Marzano model is an emphasis on assessing a teacher's use of technology.

In all of the models discussed, none account for the evaluation of music teachers. Within the instructional domain of the models in particular, the degree to which each is conducive to the evaluation of the music teacher is by virtue of what each directs the evaluator not to observe or consider in relation to traditional classroom practice. Of the four models discussed here, the Marzano and Danielson models have been found to be

the most classroom specific, followed by McREL, then by Stronge as being the least oriented to a particular instructional environment.

Chapter 3

Methodology

Introduction

The purpose of this research was to examine the perceptions of New Jersey public school music teachers regarding the evaluations they received as mandated by state policy. At the time of this study, New Jersey public school music teachers were evaluated through the use of an observational tool, a lens through which an evaluator, regardless of content expertise, was directed in the observation of predetermined competencies. By 2015, the New Jersey Department of Education had approved 23 possible models from which a district could choose (New Jersey Department of Education, 2015b). Of those, two of the most used and most contrasting models have been employed in this study.

A public school music teacher in New Jersey could be subject to at least four combinations of administrator and model types depending on where they teach: 1) a content-knowledge administrator using evaluation model “A”, 2) a non-content knowledge administrator using evaluation model “A”, 3) a content-knowledge administrator using evaluation model “B”, or 4) a non-content administrator using evaluation model “B”. Each scenario offers the potential for dissimilar outcomes in influencing perceptions and attitudes. The potential for dissimilar outcomes provided the rationale for this investigation. In order to determine whether or not dissimilar perceptions existed on account of differing evaluation conditions among New Jersey’s public school music teachers, the following research questions were formulated:

1. Will administrator type (administrator with a degree in music or not) have a significant effect upon music teacher perception of the evaluation process in New Jersey?

2. Will evaluation model type (Danielson FfT or Stronge TEPES) have a significant effect upon music teacher perception of the evaluation process in New Jersey?

3. Will administrator and evaluation model type interaction have a significant effect upon music teacher perception of the evaluation process in New Jersey?

Population

The participants of this study were limited to New Jersey music teachers possessing K-12 music teacher certification, and who had at least a half-time teaching assignment or more within a New Jersey public school. This delimitation was necessary to insure that the population was made up of teachers who were evaluated by administrators either possessing content specialty or not, and who were evaluated by one of New Jersey's approved evaluation models. The author chose the state of New Jersey because of his familiarity with the region as well as its state-approved evaluation models. Of greater significance, of the states which have been examined previously in this area of research, New Jersey had yet to be fully examined in a state-wide study.

The targeted population included music teachers who taught within all areas that New Jersey public schools offered during the 2015-2016 academic year. Access to this population as a whole was made primarily through a statewide association of public school music teachers known as the New Jersey Music Educators Association (NJMEA) whose total "Full Active" membership at the time of this research was known to be 1,981 members (W. McDevitt, personal communication, April 21, 2016). In addition, music teachers evaluated under the Stronge Teacher Effectiveness Performance Evaluation System were solicited directly via mailed surveys.

Participation within this research was entirely voluntary with no inducement or incentive offered to participants other than that findings would be made available via the NJMEA member webpage *Tempo Express* at the conclusion of the research.

Instrumentation

In order to investigate the attitudes and perceptions of this population, survey methodology was utilized that employed quantitative elements. The survey instrument for this study was created by the author and was informed from the framework of *Elements and Related Themes* (see Table 2) that resulted from the author's synthesis of the relevant literature on music teacher evaluation in general and for New Jersey in particular. The four major areas within this framework consisting of *Teacher*, *Evaluator*, *Evaluation*, and *Result* found parallel in the survey's makeup of three Domains consisting of 1) Teacher, 2) Evaluation Model and Evaluator Type, and 3) Resulting Usefulness and Professional Growth. What resulted was a five-point Likert Scale survey delineated into three major considerations followed by demographic questions, concluding with space made available for an open-ended verbal response (see Appendix A). Each domain is described below.

Domain 1, "You, the Teacher," asked participants to indicate the degree to which they believed their last evaluation accurately assessed their own performance as a teacher during their last lesson or rehearsal. Areas included instructional and rehearsal concerns more specific to the music environment than otherwise found in the non-music classroom in addition to general classroom instructional areas.

Domain 2, "Evaluator and Evaluation Instrument" directed participants to consider the qualifications, appropriateness, and effectiveness of their particular evaluator

as well as the suitability and appropriateness of the evaluation model used to evaluate them. This domain combined the separate elements of *Evaluator* and *Evaluation* into one whole to better streamline the instrument.

Domain 3, “Resulting Usefulness and Professional Growth” asked participants to gauge to what extent the finished evaluation was useful in assisting their future growth, if participants would recommend the same evaluator in the evaluation of a fellow music teacher, and if they would look forward to another evaluation conducted in similar circumstances as an opportunity for further professional growth. Domains 1, 2, and 3 in all contained a total of 18 dependent measure questions, each with a 1 to 5 Likert Scale rating.

Following was a section for demographic questions that were informed from prior studies examining music teacher evaluation perception (Becher, 2011; Maranzano, 2002). Demographics were obtained in the areas of number of years teaching, grade levels taught, area of specialization, standardized test score data tied to evaluation, merit pay, school population, number of students taught weekly, administrator background, and NJMEA region. In order to investigate New Jersey’s particular teacher evaluation environment, a list of the New Jersey Department of Education’s 23 state-approved evaluation models for 2015-2016 was included from which participants could select. In all, demographics were limited to 13 questions. Space was made available for an open-ended verbal response, inviting participants to supply a written comment about their perception of music teacher evaluation in their district.

Once created, the instrument was further revised following suggestions provided by the dissertation committee. The final draft of the survey contained a total of 31

questions. It was estimated that the average time needed to complete the survey would be five minutes, or just under 10 seconds per question excluding additional time needed for the open-ended verbal response.

The survey was inputted electronically into a SurveyMonkey platform containing three provisions in order to protect participants: 1) Secure Sockets Layer (SSL) protection was enabled so that connections between participants and the server was secure, 2) Internet Protocol (IP) address tracking was disabled making respondents anonymous, and 3) an informed consent form was included on page one of the survey (see Appendix B).

Data collection. The survey was distributed in three forms: 1) internet link, 2) in-person distribution of physical surveys, and 3) mailing.

Internet link. Following committee and IRB approval (Appendix J) and input of the survey electronically, the first access of the survey occurred on January 5, 2016 to the membership of the New Jersey Music Teachers Association (NJMEA) via its member page *Tempo Express*. NJMEA President, William McDevitt had agreed in the months leading up to survey's launch to allow access to it through the association's website. Following the launch, a second email occurred on February 4.

In-person distribution. Following the first access to the survey being made electronically on January 5, the next dissemination of the survey occurred 14 days later as a hard-copy version that was distributed directly to New Jersey public school music teachers who attended the 2016 NJMEA State Conference on February 18-20 at the East Brunswick Hilton, East Brunswick, New Jersey. This annual event draws music educators for its professional development offerings and student showcases. Six hundred-fifty hard-copy versions of the survey identical to the online version were created for this

purpose, this number based on conference registration data. The hard-copy version of the survey consisted of four pages formed from one 11x17 inch sheet folded in half, the first page of which containing introductory material, participant information, and informed consent form with the three remaining pages containing the survey questions themselves. The introductory material informed participants that they could participate in the research in three ways by 1) completing and returning the physical survey during the convention, or 2) accessing the survey via the *Tempo Express* email (*All-State Band*, February 4, 2016) notifying NJMEA members of the research opportunity that included a link to the survey, or 3) going directly to the SurveyMonkey link stated on the form (see Appendix C).

The survey was distributed and collected by hand to individual convention participants from 7:00 a.m. to 4:30 p.m. on day one of the convention, and from 7:30 p.m. to 4:30 p.m. on day two. Before distributing each survey, it needed to be determined that each met participant qualifications for the research. On day three of the convention, surveys were neither handed out nor physically collected. A notice was included on the survey that if participants were unable to return completed surveys, they could transfer their responses to the online format.

Distribution and collection of the survey was facilitated by the author and two assistants. The two assistants were previously informed about the nature of the study and were instructed in proper survey protocol and security. Author and assistants wore t-shirts created specifically for the purpose of identifying persons involved in the distribution and collection of surveys (see Appendix D). Author and assistants formed a large triangle in the lobby of the hotel lobby with one assistant taking his/her place on the balcony

overlooking the main lobby (this balcony being a main concourse to and from convention offerings). Each member of the distribution/collection team wore a secure tote bag over the shoulder for the collection of responses. Completed surveys were removed at intervals from the bags and were secured under lock and key. By the end of day two, all 650 surveys had been distributed. Following the convention, returned surveys were manually entered into the SurveyMonkey platform under a separate collector labeled as *Convention*. Returned surveys were referenced by date collected with a corresponding reference made to each SurveyMonkey entry.

Mailing. Before the initial launch of the survey, it was realized from a statistical standpoint that an additional step of mailed surveys might be necessary in the event that the number of responses collected electronically and at the NJMEA convention from teachers active in Stronge districts would prove to be insufficient. It was determined that at least 25 responses each come from teachers who were evaluated by content and non-content specialists active within Stronge districts. Owing to an 11 percent use of the Stronge model by New Jersey districts overall (Mooney, 2013) might make the collection of an adequate number of responses from these districts challenging. By March 1, it was determined following an examination of responses collected to that point both electronically and from the convention that additional mailed surveys to teachers in districts using the Stronge framework would be necessary.

Beginning on March 9, surveys similar to those handed out at the NJMEA convention were mailed with self-addressed stamped envelopes to music teachers employed in Stronge districts. This was accomplished through a list of New Jersey school districts in which it was determined that the Stronge Teacher Evaluation Performance

System had been adopted (Appendix E). Individual school websites were consulted for music educator names and school addresses. Envelopes included one music teacher's name per envelope and were hand addressed. In the few instances in which a school website did not provide names, these schools received at least two envelopes addressed separately to instrumental and vocal music teachers at the elementary and middle school level, and an additional letter to a possible string position at the high school level. A post office box was rented for the collection of responses. Mailed surveys to teachers ended March 14 with 380 surveys having been mailed to every current music teacher position in New Jersey where the Stronge Teacher Evaluation Performance System was identified as being employed. Returned surveys were manually entered into the SurveyMonkey platform under a separate collector labeled as *Stronge Mailing*. Returned surveys and envelopes were referenced by date collected with a corresponding reference made within each SurveyMonkey entry.

Data Analysis

Quantitative considerations. The central question of this research was to what extent the independent variables of administrator and evaluation model types have in influencing music teacher perception of the evaluation process, and to what extent these variables interact. To that end, a two-way ANOVA factorial design was utilized. The design examined 1) a main effect for administrator type (administrator with a degree in music or not), 2) a main effect for evaluation model type (Danielson Framework for Teaching or Stronge Teacher Effectiveness Performance Evaluation System), and 3) an interaction effect between administrator and model types. As is the case with such a design, three null hypotheses resulted:

1. H_0 : Administrator type will have no significant effect upon music teacher perception of the evaluation process.
2. H_0 : Evaluation model type will have no significant effect upon music teacher perception of the evaluation process.
3. H_0 : Administrator and evaluation model type interaction will have no significant effect upon music teacher perception of the evaluation process.

The strength of evidence required to either reject or not reject the null hypotheses was intended to come from ANOVA results generated from a random sample of no less than 100 responses each coming from music teachers in Danielson and Stronge districts, with no less than 25 of those responses each coming from teachers evaluated by an administrator possessing a degree in music. The Alpha level for significance was set at $p < .05$ for significant and $p < .01$ for highly significant.

Following the collection and categorization of responses according to independent variables, each of the 18 dependent variables of participant attitudes and perceptions were examined for significant differences (evidence against the null hypotheses, if any) within the two main effects for administrator and model types, and interaction between them.

Qualitative considerations. A qualitative element was included in the form of an open-ended response concluding the survey (see Appendix A). The nature of the question was designed so as not to direct participants in the consideration of any one aspect or aspects of music teacher evaluation nor were they directed regarding the length of response.

The phenomenological approach was determined to be best in considering participant responses in that it was deemed important “to understand several individuals’ common or shared experiences of a phenomenon” (Creswell, 2007, p. 60).

Following the collection of responses, significant statements, sentences, or quotes were highlighted that provided an understanding of how participants experienced music teacher evaluation. Next, “clusters of meaning” from these significant statements were developed into themes (p. 61). Themes were then composited into one of the four evaluative conditions depending on the respondent (Danielson/administrator with music degree; Danielson/administrator without music degree; Stronge/administrator with music degree; Stronge/administrator without music degree). Composites were then examined for frequency of appearance for each identified theme within each condition and listed in thematic tables and figures. Findings were then examined independently and in comparison with quantitative results for relationships to determine to what extent qualitative and quantitative findings corroborated or not.

Chapter 4

Results

Introduction

The purpose of this study was to investigate New Jersey public school music teacher perception of the evaluation process with regard to the potential interaction of administrator type with evaluation model type in influencing perception. The independent variable of administrator type was determined by whether the evaluator conducting the evaluation possessed a degree in music or not. The independent variable of evaluation model type was determined by whether the teacher was evaluated by either the Danielson Framework for Teaching (Danielson) or the Stronge Teacher Effectiveness Performance Evaluation System (Stronge). Towards that end, a survey was developed for this purpose, its content being informed collectively from the objectives of this research as well as from the literature on music teacher evaluation research. The finished survey (see Appendix A) was ultimately disseminated in three phases, each identified by a primary modus of response:

Phase 1- electronic mailing to the total active membership ($N = 1981$) of the New Jersey Music Educators Association (NJMEA) on January 5 and February 4, 2016.

Phase 2 - distribution and collection of the survey in hard-copy format at the 2016 NJMEA State Conference on February 18 and 19 with the option of responding on-line.

Phase 3 - mailing and return via return envelope of the hard copy survey to teachers within districts determined to utilize the Stronge Teacher Evaluation Performance System beginning on March 9 with the option of responding on-line.

Phase 3 of the survey had been provided for and was held in reserve in the event that the number of Stronge responses collected from phases 1 and 2 had shown to be inadequate. For statistical purposes, it was initially determined that a minimum of 100 responses each come from both Danielson and Stronge districts with at least 25 each coming from teachers evaluated by an administrator possessing a degree in music. By March 1, with a total of 36 Stronge responses having been collected at that point from phases 1 and 2, Phase 3 proved necessary. Beginning on March 9, Phase 3 provided for 380 mailed surveys to individual music teachers determined to have been active within Stronge districts (see Appendix E). The survey was formally closed on March 31.

Survey Return Rate and Origin

By April 15, a total of 418 responses had been received. Of those, 414 were from members of the NJMEA representing 21 percent of NJMEA's total active membership ($N = 1981$) during the time the survey was active from January to the end of March, 2016. Three hundred eighty-eight of the collected surveys were determined to have been complete. The 30 responses not included in the sample were found to be either 1) partially complete ($n = 11$), 2) failed to identify the evaluation model used in the respondent's district ($n = 17$), or 3) did not indicate a specific administrator type ($n = 2$). Table 7 details the completed survey returns for all three phases along with method of return for all respondents inclusive of all evaluation models reported.

Table 7

Survey Returns by Method of Return and Response Rate

	<i>Tempo Express</i> Electronic Mailing	NJMEA Convention	Mailing to Stronge Districts	Total
Online responses	169	13	23	205
Handed-in responses	–	124	–	124
Mailed responses	–	–	59	59
Total	169	137	82	388
Response rate	9%	21%	22%	20%

Note. A dash indicates no response.

One of the purposes of this study having been to examine the interaction between two particular evaluation models and evaluator types, the total sample was then examined for responses originating from Danielson and Stronge districts. Of the original sample, 308 were found to come from districts that used one of these two models. It is noted that Phase 3 of the survey (Stronge mailing) contributed six additional Danielson responses owing to a district (or districts) that were initially considered to have been Stronge districts. Table 8 details the survey's three phases for all Danielson and Stronge responses along with method of return.

Table 8

Survey Returns from Danielson and Stronge Districts by Method of Return

	Danielson Responses				Stronge Responses			
	<i>Tempo Express</i> electronic mailing	NJMEA convention	Stronge Mailing	Total	<i>Tempo Express</i> electronic mailing	NJMEA convention	Stronge mailing	Total
Online responses	108	6	2	116	19	–	26	45
Handed-in responses	–	73	–	73	–	17	–	17
Mailed responses	–	–	4	4	–	–	53	53
Total	108	79	6	193	19	17	79	115

Note. A dash indicates no response

This sample was further examined for which Danielson and Stronge respondents were evaluated by an administrator either possessing a degree in music or not. Each response from the Danielson and Stronge sample ($n = 308$) fell within one of four possible categories or treatments as shown in Table 9.

Table 9

Survey Returns from Danielson and Stronge Respondents by Administrator Type

	Danielson Districts	Stronge Districts	Total
Administrator with music degree	47	28	75 (24%)
Administrator without music degree	146	87	233 (76%)
Total	193 (63%)	115 (37%)	308

The remaining responses ($n = 80$) received from teachers in districts that utilized models other than those by Danielson and Stronge are listed in Appendix F. Also included in this appendix are demographic data for survey questions not covered in this chapter for Danielson and Stronge respondents.

ANOVA Findings Overview

The Likert scale responses taken from the survey from music teachers evaluated by either the Danielson or Stronge models were subjected to a two-way between-subjects analysis of variance having two levels of administrator type (administrators who possessed a degree in music and those who did not) and two levels of evaluation model (Danielson and Stronge). Statistical significance was set *a priori* at $p < .05$ for significant and $p < .01$ for highly significant.

Before presenting the findings of the dependent variables separately, an overview is shown in Tables 10, 11, and 12. The overview includes results for the factors of administrator and model type as well as the interaction between them. Questions have been numbered to facilitate their reporting. If no significance was found within a question's main effects of administrator and evaluation model types, then the direction of difference (in this case, which condition was higher than the other) was not reported. Attention will be drawn to significant differences only having been found in Domains 2 and 3 from the sample obtained for this research.

Table 10

Overview of Results for Domain 1, “You the Teacher”

	Survey Question	Administrator Type	Direction of Difference	Model Type	Direction of Difference	Administrator/ Model Interaction
1.1	Your ability to hold the interest of the class/rehearsal	ns	—	ns	—	ns
1.2	Your effectiveness using verbal communication	ns	—	ns	—	ns
1.3	Your effectiveness using non-verbal communication	ns	—	ns	—	ns
1.4	Your ability to redirect off-task behavior	ns	—	ns	—	ns
1.5	Your ability to articulate learning/artistic objectives	ns	—	ns	—	ns
1.6	Your ability to motivate students towards learning/artistic objectives	ns	—	ns	—	ns
1.7	The consideration of abilities unique to your own teaching/rehearsal style	ns	—	ns	—	ns

Note. “ns” indicates no significant difference. A dash indicates no direction is reported. Summary statistics are found in Appendix G.

ns $p > .05$

Table 11

Overview of Results for Domain 2, “Evaluator and Evaluation Instrument”

		Administrator Type	Direction of Difference	Model Type	Direction of Difference	Administrator/ Model Interaction
2.1	The administrator was qualified to evaluate you	**	Degree	ns	—	ns
2.2	The person conducting the evaluation makes a meaningful contribution to your musical development and growth	**	Degree	ns	—	ns
2.3	Your professional relationship with the administrator is positive	ns	—	ns	—	ns
2.4	The evaluation instrument used to evaluate you was appropriate for music	ns	—	ns	—	ns
2.5	The evaluation instrument used to evaluate you was suited to your area of specialization	*	Degree	ns	—	ns
2.6	The evaluation instrument your district uses best reflects your own evaluation philosophy	ns	—	ns	—	*
2.7	You would welcome the administrator to evaluation you again if you had a choice	*	Degree	ns	—	ns

Note. “Degree” indicates administrator with a music degree. “ns” indicates no significant difference. Asterisks indicate significant differences. Summary statistics are found in Appendix H.

ns $p > .05$ * $p < .05$ ** $p < .01$

Table 12

Overview of Results for Domain 3, “Usefulness and Professional Growth”

		Administrator Type	Direction of Difference	Model Type	Direction of Difference	Administrator/ Model Interaction
3.1	You found the evaluation useful towards your professional growth	ns	—	*	Danielson	**
3.2	You incorporated the suggestions from your evaluation into your future lessons/rehearsals	ns	—	*	Danielson	ns
3.3	You would recommend the same administrator in the evaluation of other music teaches for their professional growth	**	Degree	*	Danielson	ns
3.4	You would look forward to the next evaluation by he same administrator as an opportunity for further professional growth	**	Degree	ns	—	ns

Note. “Degree” indicates administrator with a music degree. “Danielson” indicates Charlotte Danielson’s Framework for Teaching. Summary statistics are found in Appendix I.

“ns” indicates no significant difference. Asterisks indicate significant differences.

ns $p > .05$ * $p < .05$ ** $p < .01$

ANOVA Findings by Domain and Question Items**Domain 1**

There were no significant differences found for any of the seven questions within Domain 1, “You, the Teacher.” The main effects for administrator and model type, and interaction effect were all non-significant at the .05 level. Summary statistics for each of the seven questions in Domain 1 are found in Appendix G.

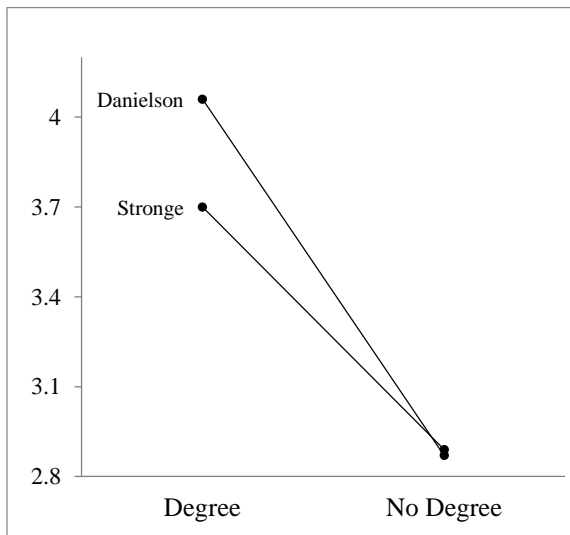
Domain 2

Significant differences were found in the main effect for degree type for five of the seven questions, and an interaction effect for one of the questions in Domain 2,

“Evaluator and Evaluation Instrument.” No significance differences were found in the main effect for model type for any of the questions.

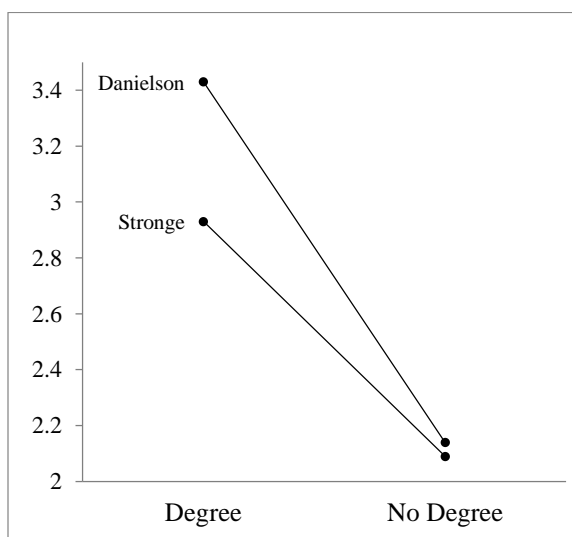
Question 2.1 “The administrator was qualified to evaluate you” yielded a main effect for degree type, $F(1, 303) = 38.50, p < .001$, such that the mean response was significantly higher for teachers who were evaluated by an administrator with a degree in music ($M = 3.93, SD = 1.14$) than for those who were not ($M = 2.88, SD = 1.17$). The main effect for model type was non-significant, $F(1, 303) = 1.13, p > .05$. The interaction effect was non-significant, $F(1, 303) = 1.34, p > .05$. Figure 1 displays the plot for Question 2.1.

Figure 1. Plot for “The administrator was qualified to evaluate you”



Question 2.2 “The person conducting the evaluation makes a meaningful contribution to your musical development and growth” yielded a main effect for degree type, $F(1, 301) = 43.73, p < .001$, such that the mean response was significantly higher for teachers who were evaluated by an administrator with a degree in music ($M = 3.25, SD = 1.32$) than for those who were not ($M = 2.13, SD = 1.10$). The main effect for model type was non-significant, $F(1, 301) = 3.06, p > .05$. The interaction effect was non-significant, $F(1, 301) = 2.02, p > .05$. Figure 2 displays the plot for Question 2.2.

Figure 2. Plot for “The person conducting the evaluation makes a meaningful contribution to your musical development and growth”

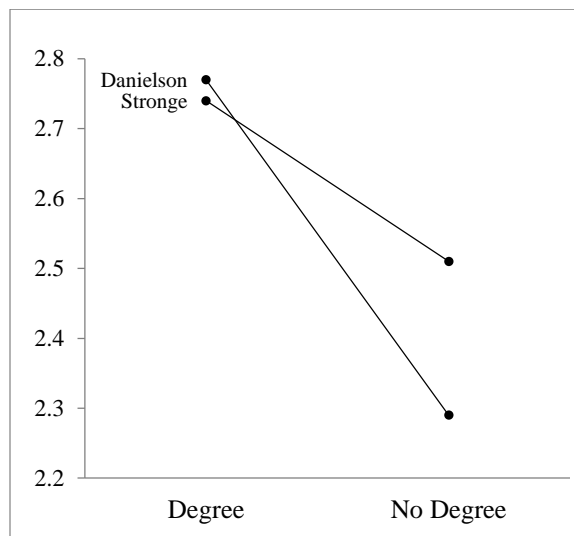


Questions 2.3 “Your professional relationship with the administrator is positive” and 2.4 “The evaluation instrument used to evaluate you was appropriate for music” did not yield significant results for main or interaction effects.

Question 2.5 “The evaluation instrument used to evaluate you was suited to your area of specialization” yielded a main effect for degree type, $F(1, 302) = 5.91, p < .05$, such that the mean response was significantly higher for teachers who were evaluated by an administrator with a degree in music ($M = 2.76, SD = 1.22$) than for those who were not ($M = 2.37, SD = 1.00$). The main effect for model type was non-significant, $F(1, 302) = 0.42, p > .05$. The interaction effect was non-significant, $F(1, 302) = 0.68, p > .05$.

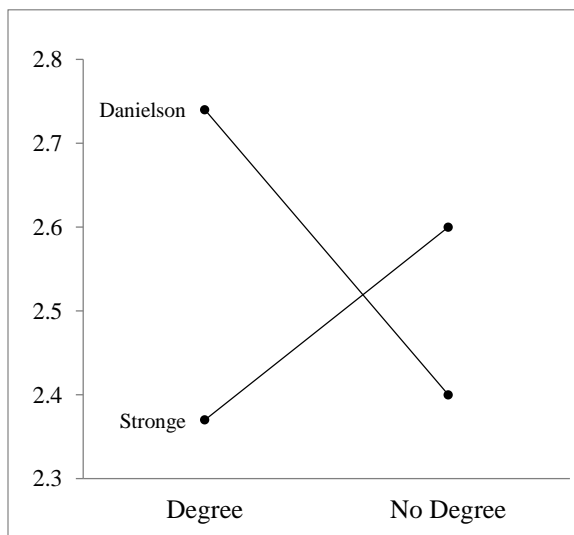
Figure 3 displays the plot for Question 2.5.

Figure 3. Plot for “The evaluation instrument used to evaluate you was suited to your area of specialization”



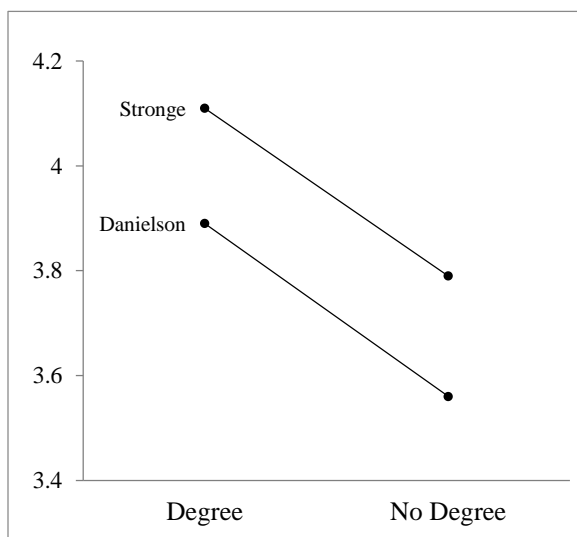
Question 2.6 “The evaluation instrument your district uses best reflects your own evaluation philosophy” yielded a main effect for degree type that was non-significant, $F(1, 303) = 0.18, p < .05$. The main effect for model type was also non-significant, $F(1, 303) = 0.38, p > .05$. However, the interaction effect was significant, $F(1, 303) = 4.12, p < .05$, indicating that the model effect was greater in the administrator with a degree condition than in the administrator without a degree condition for Danielson respondents, with an opposite trend for Stronge. Figure 4 displays the plot for Question 2.6.

Figure 4. Plot for “The evaluation instrument your district uses best reflects your own evaluation philosophy”



Question 2.7 “You would welcome the administrator to evaluate you again if you had a choice” yielded a main effect for degree type, $F(1, 303) = 4.42, p < .05$, such that the mean response was significantly higher for teachers who were evaluated by an administrator with a degree in music ($M = 3.97, SD = 1.13$) than for those who were not ($M = 3.65, SD = 1.11$). The main effect for model type was non-significant, $F(1, 303) = 2.11, p > .05$. The interaction effect was non-significant, $F(1, 303) = 0.002, p > .05$. Figure 5 displays the plot for Question 2.7.

Figure 5. Plot for “You would welcome the administrator to evaluate you again if you had a choice”



Summary statistics for each of the seven questions in Domain 2 are found in Appendix H.

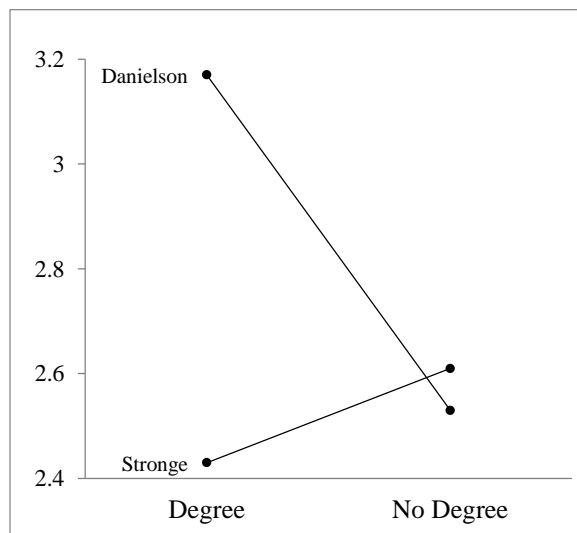
Domain 3

Significant differences were found in the main effect for degree type for two of the four questions, the main effect for model type in three of the questions, and an

interaction effect for one of the questions in Domain 3, “Resulting Usefulness and Growth.”

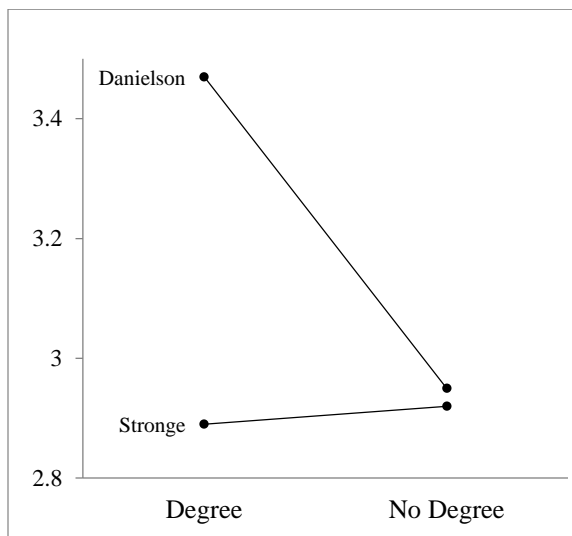
Question 3.1 “You found the evaluation useful towards your professional growth” yielded a main effect for degree type that was non-significant, $F(1, 304) = 2.43, p > .05$. However, the main effect for model type yielded an F ratio of $F(1, 304) = 4.95, p < .05$ indicating that the mean response was significantly higher for those evaluated with the Danielson model ($M = 2.68, SD = 1.11$) than those with the Stronge model ($M = 2.57, SD = 1.07$). The interaction effect was highly significant, $F(1, 304) = 7.72, p < .01$, indicating that the model effect was greater in the administrator with a degree condition than in the administrator without a degree condition for Danielson respondents, with an opposite trend for Stronge. Figure 6 displays the plot for Question 3.1.

Figure 6. Plot for “You found the evaluation useful towards your professional growth”



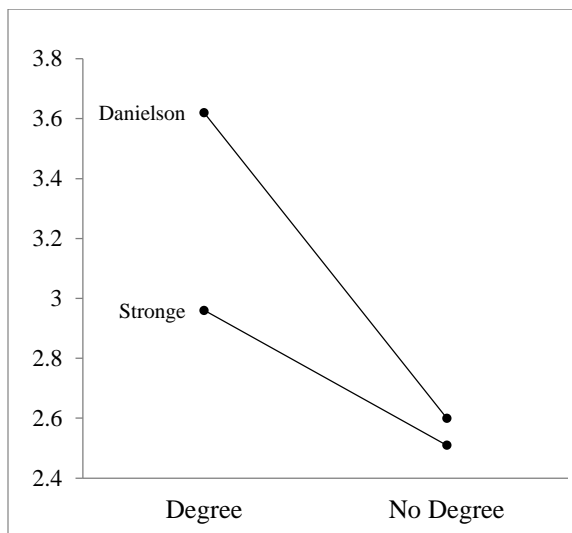
Question 3.2 “You incorporated the suggestions from your evaluation into your future lessons/rehearsals” yielded a main effect for degree type that was non-significant, $F(1, 304) = 2.71, p > .05$. However, the main effect for model type yielded an F ratio of $F(1, 304) = 3.98, p < .05$ indicating that the mean response was significantly higher for those evaluated with the Danielson model ($M = 3.07, SD = 1.09$) than those with the Stronge model ($M = 2.91, SD = 1.12$). The interaction effect was non-significant, $F(1, 304) = 3.33, p > .05$. Figure 7 displays the plot for Question 3.2.

Figure 7. Plot for “You incorporated the suggestions from your evaluation into your future lessons/rehearsals”



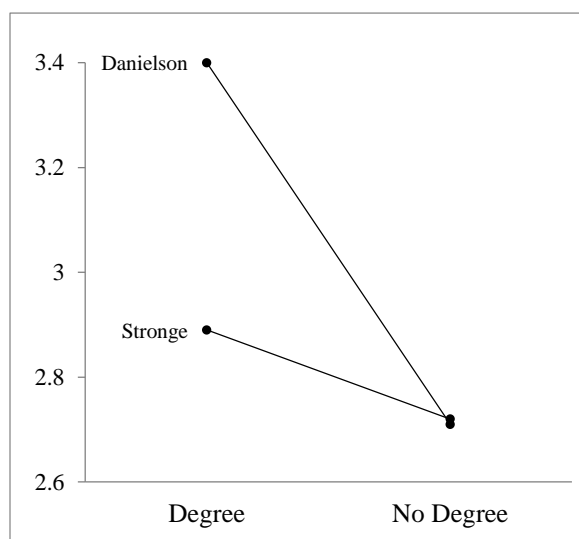
Question 3.3 “You would recommend the same administrator in the evaluation of other music teachers for their professional growth” yielded a main effect for degree type, $F(1, 304) = 21.19, p < .001$, such that the mean response was significantly higher for teachers who were evaluated by an administrator with a degree in music ($M = 3.37, SD = 1.31$) than for those who were not ($M = 2.56, SD = 1.13$). The main effect for model type yielded an F ratio of $F(1, 304) = 5.34, p < .05$ indicating that the mean response was significantly higher for those evaluated with the Danielson model ($M = 2.84, SD = 1.22$) than those with the Stronge model ($M = 2.62, SD = 1.22$). The interaction effect was non-significant, $F(1, 304) = 3.06, p > .05$. Figure 8 displays the plot for Question 3.3.

Figure 8. Plot for “You would recommend the same administrator in the evaluation of other music teachers for their professional growth”



Question 3.4 “You would look forward to the next evaluation by the same administrator as an opportunity for further professional growth” yielded a main effect for degree type, $F(1, 304) = 7.23, p < .01$, such that the mean response was significantly higher for teachers who were evaluated by an administrator with a degree in music ($M = 3.21, SD = 1.25$) than for those who were not ($M = 2.72, SD = 1.14$). The main effect for model type was non-significant, $F(1, 304) = 2.44, p > .05$. The interaction effect was non-significant, $F(1, 304) = 2.67, p > .05$. Figure 9 displays the plot for Question 3.4.

Figure 9. Plot for “You would look forward to the next evaluation by the same administrator as an opportunity for further professional growth”



Summary statistics for each of the four questions in Domain 3 are found in Appendix I.

Open-ended Response

A provision was made in the survey for an open-ended response (see Appendix A). Table 13 compares the ratio of the total number of surveys received from Danielson

and Stronge participants with the number of surveys in which participants left an open-ended response. Attention is drawn to the following: overall, roughly half of the respondents (47 percent) left an open-ended response. Respondents from Danielson districts evaluated by an administrator with a music degree responded the least (38 percent) while those in Danielson districts evaluated by someone without a music degree responded the most (51 percent). Overall, respondents in districts in which an administrator did not possess a music degree responded more (48 percent) than those who did (41 percent).

Table 13

Open-ended Response Return Rate

	Danielson			Stronge			Total surveys	Total responses	Ratio
	Surveys	Responses	Ratio	Surveys	Responses	Ratio			
Degree	47	18	38%	28	13	46%	75	31	41%
No Degree	146	75	51%	87	38	44%	233	113	48%
Total	193	93	48%	115	51	44%	308	144	47%

Following the examination of the responses for meanings and themes, they were composited according to one of the four conditions depending on the evaluative condition of the respondent (Danielson/administrator with music degree; Danielson/administrator without music degree; Stronge/administrator with music degree; Stronge/administrator without music degree). Composites were then examined for frequency of appearance for each identified theme within each condition. Table 14 displays the most common identified themes of respondents from districts in which both Danielson and Stronge models were used in which the evaluator possessed a degree in music by number of

responses received. Alternatively, Table 15 displays identified themes of respondents from districts for both models in which the evaluator did not possess a degree in music. Following these two tables, Figures 10 and 11 display the percentage of occurrence for each theme within each evaluative condition.

Table 14

Themes by Frequency for Danielson and Stronge Models from Respondents Evaluated by an Administrator with a Degree in Music

	Model incompatibility with music-overt focus on classroom management considerations	Evaluation not an issue	Administrator/teacher disconnect-administrator not cognizant of current music teacher realities	Administrator adapts model accordingly	Administrator incompatibility with discreet music area being assessed	PARCC/SGO concerns	N
Danielson	7	3	2	5	1	—	18
Stronge	1	4	4	—	2	2	13
Total	8	7	6	5	3	2	31

Note. A dash indicates no response was received. PARCC is for Partnership for Assessment of Readiness for College and Careers. SGO is for Student Growth Objectives.

Table 15

Themes by Frequency for Danielson and Stronge Models from Respondents Evaluated by an Administrator without a Degree in Music

	Administrator shortcoming due to lack of content knowledge	Model incompatibility for music—overt focus on classroom management considerations	Absence of meaningful feedback towards professional growth	Evaluation not an issue	Administrator limiting observations to classroom settings only	PARCC/SGO concerns	N
Danielson	28	24	8	7	6	2	75
Stronge	10	6	11	6	3	2	38
Total	38	30	19	13	9	4	113

Note. A dash indicates no response was received. PARCC is for Partnership for Assessment of Readiness for College and Careers. SGO is for Student Growth Objectives.

Figure 10. Themes by Percentage for Danielson and Stronge Models from Respondents Evaluated by an Administrator with a Degree in Music

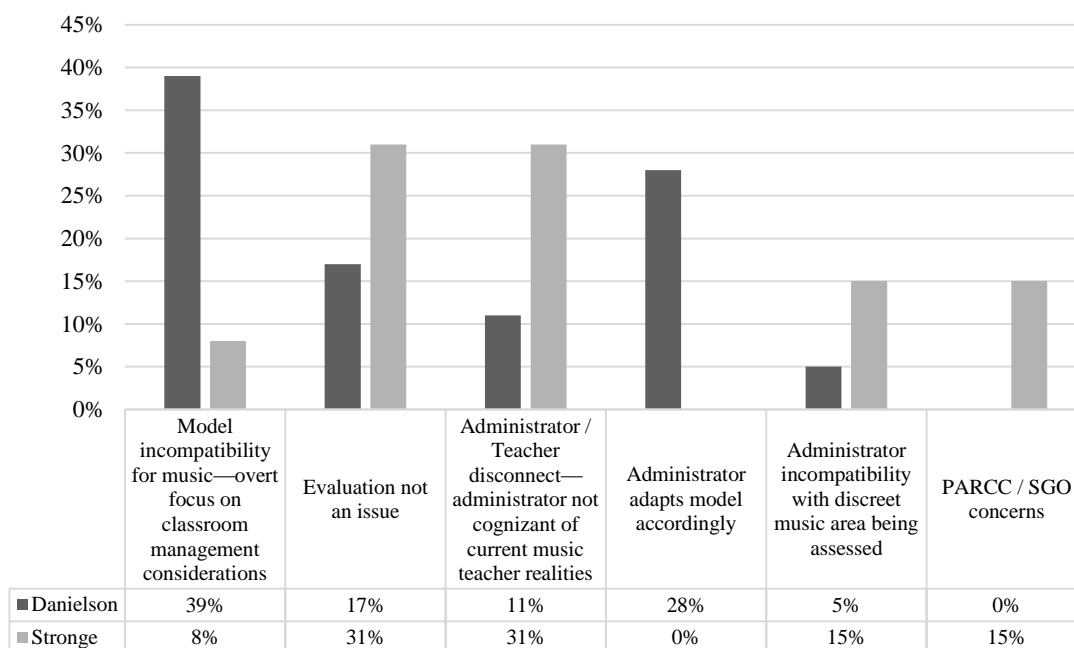
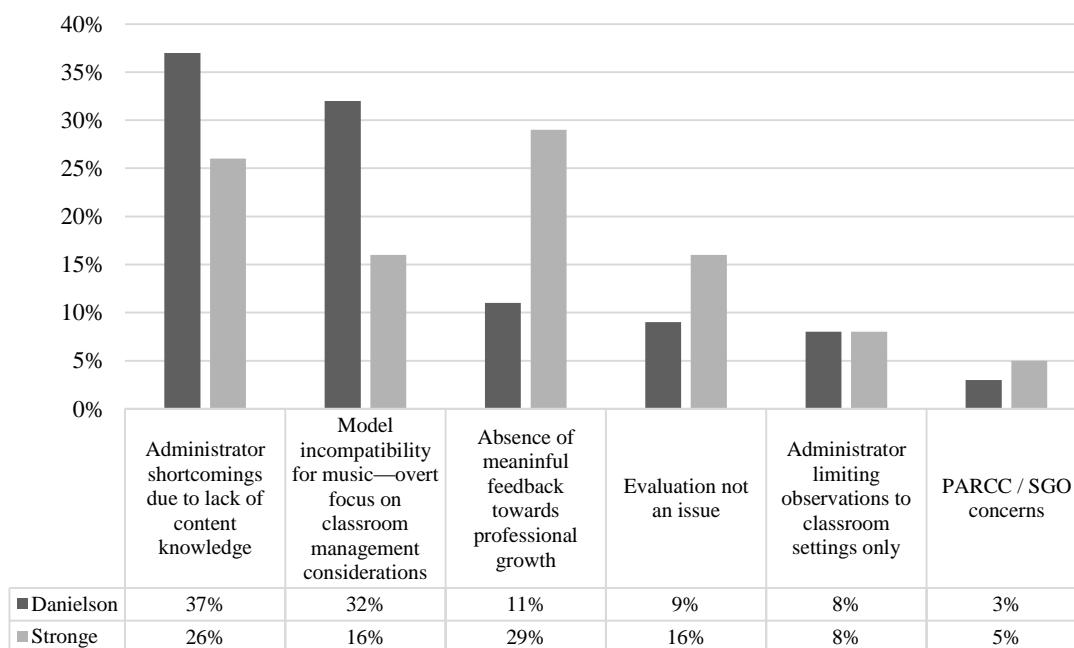


Figure 11. Themes by Percentage for Danielson and Stronge Models from Respondents Evaluated by an Administrator without a Degree in Music



Related themes were encountered between the two conditions as shown in Tables 14 and 15 and in Figures 10 and 11. Both evaluative conditions contained criticisms of evaluator and model type, concerns regarding standardized tests with specific mention of the Partnership for Assessment of Readiness for College and Careers (PARCC), and the mandated creation of Student Growth Objectives (SGO's) to measure student growth. Evaluation was also perceived as not being a concern by some participants.

Themes emerged, however, that were not shared between conditions. One was the criticism made by music teachers evaluated by administrators with a degree in music that the administrator lacked specific content knowledge in the area they were assessing. Another was the adaptation of a particular evaluation model to make it more content appropriate. Among music teachers evaluated by an administrator not possessing a degree in music was the criticism of the absence of feedback towards professional growth, and that administrators confined their evaluations exclusively to classroom settings when they had opportunity to evaluate their teachers in performance settings.

Response examples. The theme of administrator incompatibility was encountered most often among those evaluated by an administrator without a degree in music. One respondent wrote,

Evaluation is so frustrating! We have to teach our administrators (all with no music background) what we do, how it's effective, plus how it fits into the Stronge model. If we fail to bring something to their attention, it's our fault when we don't receive credit for it. (Stronge respondent evaluated by an administrator without a degree in music)

A Danielson respondent wrote, “I have never had a supervisor ‘know’ what material or who I teach.”

Model incompatibility followed administrator incompatibility among those not evaluated by someone with a degree in music. Two responses from Stronge and Danielson districts respectively ranged from “...the evaluation method our district uses forces much of music evaluation to fit into a system not written for our specialty—like fitting a round peg into a square hole” to “It’s soul killing...I found the amount (77) of things looked for so overwhelming that I am losing sleep trying to save my job.”

“Absence of meaningful feedback towards professional growth” emerged as a theme exclusively in the administrator-without-a-music-degree condition for both models. A Stronge respondent wrote, “I have received...evaluations that are ‘glowing.’ While that is nice, I would love more constructive feedback to help me grow as a music educator.” A Danielson respondent wrote, “I have never in my professional career been evaluated by a fellow music teacher. What a joy that would be; and oh how differently my teaching style might have developed if I had had that opportunity all these years.”

Criticism was not reserved for administrators lacking a music degree. Under the theme “Music teacher/administrator disconnect,” a Stronge respondent wrote, “He has a very old music degree and has been out of the classroom and any meaningful professional development activity in music in the last 15 years. He gives me positive feedback, but the feedback provides nothing to help me grow as a teacher.” A Danielson respondent wrote, “I hate to admit this, but my music supervisor is not competent to evaluate any music teacher in the district.”

Further criticism fell under the theme “Administrator incompatibility with discreet music area being assessed” in which teachers evaluated by an administrator with a music degree expressed shortcomings in their music specialists to evaluate their area of specialty. A Stronge respondent observed,

While valuable feedback is received at times, rarely is it of musical value. Our administrator is a ‘choral-guy’ and sometimes I (we) feel that he does not really understand the band side of things and the unique challenges this field presents.

(Stronge respondent evaluated by an administrator with a degree in music)

Positive positions were encountered for the evaluation process between administrative conditions. In the theme “Evaluation not an issue” for teachers evaluated by an administrator with a music degree, a Stronge respondent wrote, “Since we are mostly observed by our fine arts supervisor, I feel like we are being evaluated fairly and well.” A Danielson respondent wrote, “We have to use Danielson and having a music supervisor is a huge benefit.” A Stronge respondent evaluated in the non-degree category stated, “If you do your job correctly...the evaluations shouldn’t be an issue.” A Danielson respondent expressed, “I am fine with the way that we are evaluated in our district. I have a very supportive administration and they evaluate us fairly and even though they don’t have a music background, they trust us in our teaching abilities.”

Respondents not evaluated by an administrator with a music degree expressed frustration that their observations were limited exclusively to classroom environments—a theme which was not to emerge in the other administrator condition. A Danielson respondent wrote, “My administrator will ONLY evaluate me on general music classes [*sic*]...I wish they would evaluate me on my rehearsals and/or lessons.” A Stronge

respondent wrote, “although he [administrator] plays piano for leisure, he has chosen to observe me in my only general music period...this seems to me an asinine practice.”

A theme to emerge exclusively from Danielson respondents was of their music-degree-administrators who adapted the Danielson model for music teacher evaluation. One respondent wrote, “We are lucky to have a supervisor who fights for our rights, understands the evaluation tool doesn’t always apply to what we teach, adjusts accordingly, and makes helpful suggestions.” Another wrote, “Overall, the evaluation system...has to be adapted to fit what we do. When the arts supervisor is involved in the process, it is generally fair.” Other expressions of this adaptation of the Danielson model included administrators who, “cleverly adapted the Danielson standards for use” and have “gone above and beyond to make us ‘fit’ into the Danielson framework,” actions that were not reported by any of the Stronge respondents in adapting the Stronge model for music teacher evaluation.

PARCC and SGO concerns were encountered as a theme, but occurring with the least frequency. A Stronge respondent without a content specialist wrote, “My only concern is the use of SGO data as a reflection of my effectiveness as a music educator.” A Danielson respondent stated, “The idea that an educator’s evaluation is in part and to varying degrees based on a child’s success on a standardized test is highly offensive.”

In summary, administrator and model suitability and professional growth concerns emerged as the dominating themes among those music teachers who chose to contribute an open-ended response. Among those teachers whose administrators adapted the Danielson model for music evaluation, evaluation was then perceived as being fairer and more useful than when it was not. Yet, the presence of an administrator with a music

degree appeared to be no guarantee of administrator suitability among some respondents. SGO and standardized test concerns were minimal when compared against the other themes to emerge. Some respondents saw evaluation as not an issue.

In the open-ended responses, music teacher evaluation appeared to be useful when the perceived appropriate administrator was conducting the evaluation and when the model was effectively adapted to music. Conversely, music teacher evaluation appeared to be perceived as least useful when model and administrator unsuitability went hand in hand.

Chapter 5

Discussion

Introduction

Protagoras (500?-411 B.C.) said, “Man is the measure of all things.” An equally true statement is, “Man is the being who measures all things” (Rudhyar, 1982, p. 28). Measurement’s importance stretches from the Levitical Law of Moses— “You shall do no wrong in judgment, in measure of weight or capacity. You shall have just balances, just weights, a just ephah and a just hin” to Einstein’s famous equation in which the measure of matter is the measure of energy itself.

Measurement is the handmaid of evaluation. Glickman stated that evaluation should ultimately “put a mirror of the classroom up to the teacher” (2012 p. 256). This study endeavored to put a mirror up to the process of observation itself, its goal to determine through teacher perception how accurately music teacher performance was being measured and to what extent this measurement proved useful.

A result of Chapter 2’s review of the literature on music teacher evaluation was the synthesis of music teacher evaluation into four domains in which something (*teacher*) is being considered (*evaluator*) by a measure (*evaluation*) with a resulting consequence (*result*). From these perspectives this study examined New Jersey music teacher attitudes and perception of the evaluation process through the development of a survey designed for this purpose. The results will be discussed for each domain as they relate to previous research.

Domain 1, “You, the Teacher”

The questions contained in Domain 1 asked music teachers to indicate the degree to which they believed their last evaluation accurately assessed their own performance as a teacher in general classroom instructional areas in addition to instructional and rehearsal areas more specific to the music environment. The results indicated that regardless of administrator or model type there were no significant differences in perception among music teachers for all of the questions found within this domain, whether the question considered classroom specific or content specific areas.

Findings found within Domain 1 reflect those of Schmidt (1992) and Edgar (2012) in which interpersonal skill, rapport, and the familiarity in assessing general classroom skills in classroom settings are shared among administrators despite training and background. Taebel’s research (1990a, 1990b) discovered an explicit emphasis on presentation and questioning as the chief components normally evaluated during a lesson regardless of discipline. Participant responses in these areas here showed no significant differences of perception regardless of evaluator or model type in the assessment of the verbal cognitive mode orientation of Domain 1.

In regards to observers not trained in music, the research of Duke (1987) found that “observers may focus their attention to a much greater extent upon teacher behaviors as opposed to those of students” (p. 122). Henninger (2002) investigating in a related vein revealed similar findings. Here, the open-ended responses from the survey support these findings in which the overt focus of teacher/classroom management considerations were at the expense of music content considerations. A greater ratio of responses in the area of overt focus on classroom management considerations were encountered from those

teaching in Danielson than for Stronge districts for both administrator conditions, perhaps owing to a higher level of classroom specificity found in the Danielson model than when compared to the Stronge model. Yet owing to the inclusion of the factor for model type in the study's quantitative design, a finding was that contrasting model types, whether highly classroom specific or not, revealed no significant difference in affecting perception within this domain.

As a whole, Domain 1 received higher means overall than compared to the other domains (see Appendixes G, H, and I). Perception within this area of instruction emerged as the least unfavorable aspect of evaluation among the respondents. Some issues concerning this domain, however, were articulated in the open-ended responses as displayed in Figures 10 and 11 (Figures 10 and 11 from Chapter 4 are reproduced below for the purpose of this discussion).

Figure 10. Themes by Percentage for Danielson and Stronge Models from Respondents Evaluated by an Administrator with a Degree in Music

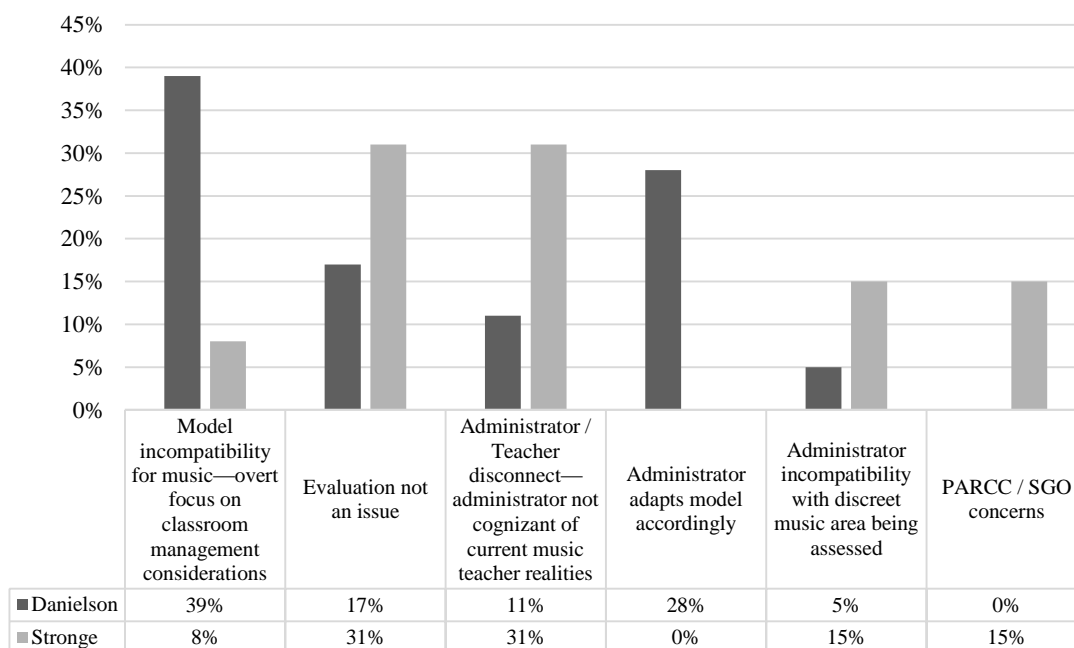
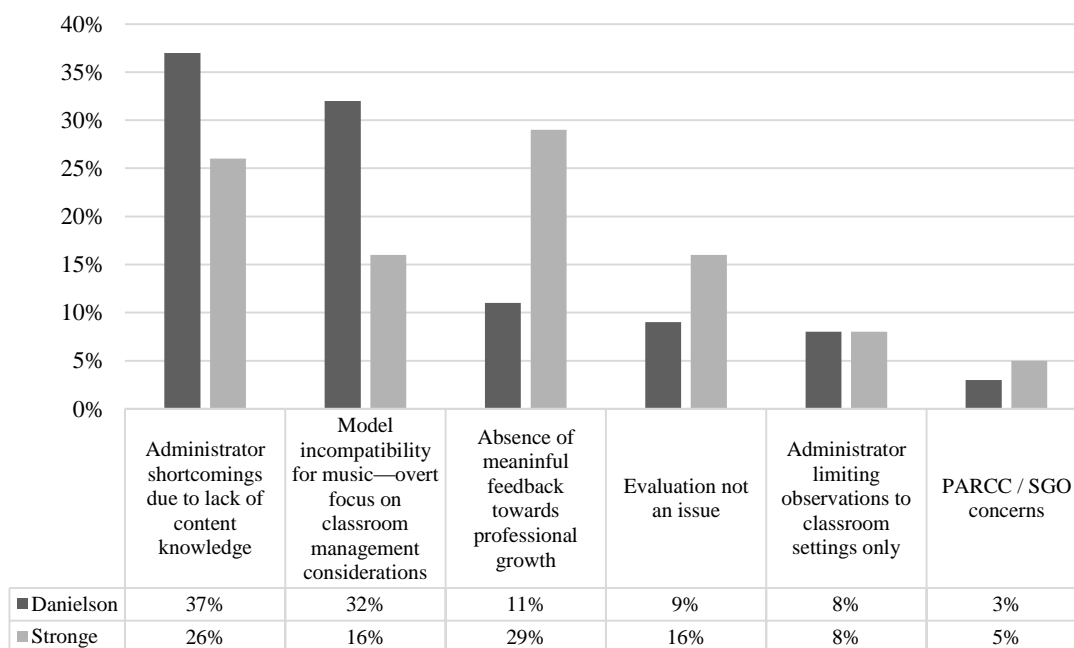


Figure 11. Themes by Percentage for Danielson and Stronge Models from Respondents Evaluated by an Administrator without a Degree in Music



A result of non-music-degree-administrator apparent preference exclusively within verbal cognitive mode classroom settings was found in the emergence of a theme identified as “Administrator limiting of observations to classroom settings only” (Figure 11). In this theme, respondents voiced shortcomings about their non-music-background administrators intentionally avoiding the consideration of respondent student performances. Such administrator behavior supports Edgar’s finding (2012) that administrators not possessing a music background may avoid considering discreet music skills in favor of assessing general classroom skills. Schmidt’s findings (1992) suggest that when administrators who are “untrained” (p.19) when forced to consider discreet musical considerations resulted in low agreement among raters. High observer agreement being a tenet of evaluation, it is not known explicitly from this research if participant administrators in this matter made their choice unilaterally in limiting evaluations to classroom settings, or if they were directed in this matter from their districts or from the creators of the evaluation models themselves.

Domain 2, “Evaluator and Evaluation Instrument”

The questions contained in Domain 2 asked music teachers to indicate the degree to which they agreed with statements pertaining to administrator and model types. As noted earlier, this section of the survey was the streamlining of the two separate domains of *evaluator*—who is performing the evaluation, and *evaluation*—the measure used, into one survey domain.

Pertaining to evaluator type, Domain 2 provided for four questions. Pertaining to model type, Domain 2 provided for three questions. For the administrator type questions, significant differences were found for the factor of administrator type for three of the four

questions. For the model type questions, significant differences were found for the factor of administrator type for one of the three questions. A significant interaction effect between administrator and model factors was found for one of the questions pertaining to model type.

Administrator Type. As was seen in Chapter 2, the literature on music teacher evaluation places *evaluator* concerns ahead of any other single issue. The results found in the present study for music teachers in New Jersey have reconfirmed this importance. Six of the eleven significant differences encountered in the ANOVA results were for the factor of administrator type. Qualitatively, respondents who contributed an open-ended response evaluated by an administrator without a degree in music articulated administrator suitability above any other to emerge as an identifiable theme (Figure 11).

The results for Question 2.1 “The administrator was qualified to evaluate you” for the factor of degree type was highly significant at the .01 level. Likewise were the findings for Question 2.2 “The person conducting the evaluation makes a meaningful contribution to your musical development and growth.” Question 2.7 in which the respondent was asked if they would welcome the same administrator to evaluate them again if they had a choice was found to be significant for administrator type. These results are identical to those of past investigations and commentary in which administrator suitability had emerged as the number one area of concern (Becher, 2011; Collins, 1996; Cope, 2003; Goddard, 2004; Maranzano, 2002; Taebel, 1990a).

It was found in this study that an administrator possessing a degree in music was no guarantee of respondent evaluative satisfaction, however. Here, Scott Shuler’s position (1996) that many arts administrators will lack content expertise in at least one of

the arts areas in which they will evaluate was brought into relief in the emergence of the open-ended response theme “Administrator incompatibility with discreet music area being assessed” (Figure 10). In several instances, respondents who were evaluated by an arts administrator possessing a degree in music voiced dissatisfaction with the evaluator. A respondent assessed their evaluations as being “rarely of musical value” on account of the evaluator being “a choral guy” and the respondent being a band teacher. Alternatively, another respondent wrote, “My evaluator is a former band director so I trust him but not totally. He...doesn’t understand the difficulties teaching in elementary.” Shuler mentions the existence of programs that provide training for the music administrator who must assess multiple fine arts content areas (1996). Perhaps the exchange of positions between the two administrators or teachers cited above would be the easier solution. In any case, with some arts administrators being responsible for music, art, dance, and even theater simultaneously, should one administrator do it all? And what of appropriate instrumentation?

Model type. Following administrator suitability in the literature is evaluation model suitability as the second most area of concern. The results found in this study examining music teacher perception in New Jersey among Danielson and Stronge respondents showed no change in this ranking. Three of the eleven significant differences encountered in the ANOVA results were for the factor of model type. Qualitatively in the open-ended responses, respondents who were evaluated by an administrator without a degree in music articulated model suitability as the second most encountered thematic element following administrator suitability (Figure 11). Respondents evaluated by

administrators already in possession of a music degree articulated model suitability as their number one concern (Figure 10).

These results confirm the earlier findings and positions regarding instrumentation's number two position (Brophy, 1993; Collins, 1996; Goddard, 2004; Maranzano, 2002; Taebel, 1990a). NAFME's own "Teacher Evaluation Position Statement" (2012) likewise placed instrumentation as its number two area of concern. New to music teacher evaluation research was this study's comparison of two contrasting evaluation models between respondents who were either evaluated by the Danielson model or the Stronge model. Both are identifiable as generic models, yet they contrast due to the comparatively higher classroom specificity of one (Danielson) to the intentionally lesser specificity of the other (Stronge).

Although Question 2.4 "The evaluation instrument used to evaluate you was appropriate for music" resulted in no significant differences for evaluator type, model type or interaction, the survey's following question "The evaluation instrument used to evaluate you was suited to your area of specialization" resulted in a significant difference for the factor of evaluator type or who was conducting the evaluation.

To arrive at a possible understanding as to why the nature of the evaluator affected the respondents' perception for a question regarding model type, in this case a question directed towards a model's ability to consider a respondent's area of specialization, and not just music in general (as was the case in Question 2.4), the open-ended responses provided some possible clues. Among Danielson respondents who were evaluated by an administrator with a degree, a theme emerged in which administrators were praised for adapting the Danielson model according to particular situation and need.

One respondent describing his/her supervisor wrote, “He has gone above and beyond [to] make us ‘fit’ into the Danielson framework, either by thinking creatively or using N/A.”

No such mention was encountered for any of the Stronge respondents.

It should be noted that in the open-ended responses, respondents were not prompted in the direction of a particular topic in which to consider. Yet the presence of the responses in which content area administrators facilitated a model to fit within a particular evaluation scenario corroborated with the statistical findings. Based on these findings, it may be that a model’s ability to account for content specific considerations was of greater importance to the respondents than for any model’s ability to assess music overall as a subject area, as was expressed in Question 2.4 “The evaluation instrument used to evaluate you was appropriate for music” in which no significant differences were found for either administrator or degree type. But key perhaps was not the nature of the instrument used in its original form, but in the ability of an administrator to modify the instrument “creatively,” “cleverly,” to make it “fit,” as described by the respondents above according to the specific need at hand. Key also would be an administrative free hand in which to do this.

The first encounter of significant interaction between administrator and model type occurred in Domain 2 for Question 2.6 “The evaluation instrument your district uses best reflects your own evaluation philosophy.” The nature of this question did not concern instructional focus as did the previous question dealing with specialization, but with a more nebulous consideration dealing with overall evaluative philosophy as it pertained to the use of the Danielson and Stronge models. In order to better consider the

result for this question, the result will be compared to those of the six other questions found in Domain 2.

When examining the direction of difference between degree conditions for both the Danielson and Stronge models for Domain 2, it was found for five of the seven questions that the cell means were lower for those evaluated by an administrator without a degree in music. In other words, the perception was less favorable when someone without a degree in music conducted the evaluation, a finding which was in accord with previous research. Yet not in accord with previous research was the finding for Question 2.6 “The evaluation instrument your district uses best reflects your own evaluation philosophy.” Whereas the direction of difference for those evaluated by Danielson did reflect the same pattern found in the results for 2.1, 2.2, 2.4, and 2.7 (see Appendix H) showing more favorable perception when the person conducting the evaluation had a degree, the Stronge respondents responded in the opposite direction—the perception was less favorable when the evaluator possessed a degree in music. The interaction that resulted from these opposite perceptions was strong enough to be considered significant at the .05 level.

The divergent result found in Question 2.6 has been examined and put into context against the other findings of Domain 2. Before an attempt is made at explaining these differences, it will be better considered against the results of Domain 3.

Domain 3 “Resulting Usefulness and Professional Growth”

The questions contained in Domain 3 asked music teachers to indicate the degree to which they agreed with statements pertaining to their evaluation’s overall contribution

to their professional growth. The results of Domain 3 will be briefly restated before discussion.

The majority of the significant findings of this study were found in this domain alone. They included two for administrator type and three for model type, in addition to one interaction effect. Qualitatively, respondents evaluated by an administrator without a degree expressed lack of feedback towards professional growth specifically as a theme to emerge in the open-ended responses (see Figures 10 and 11).

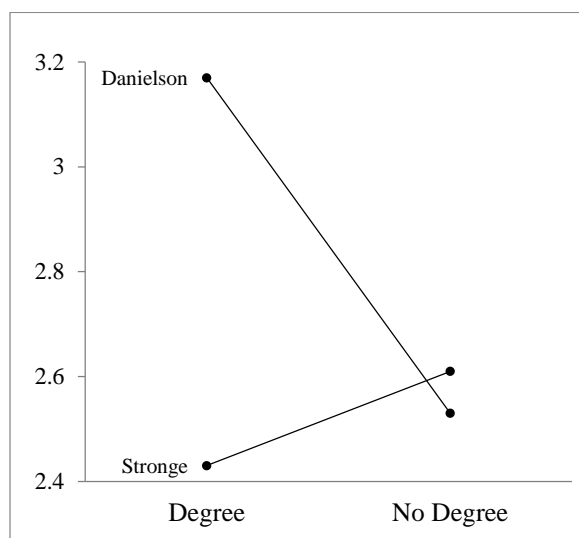
Lack of professional growth due to administrator and evaluation model shortcomings has been investigated in mixed-methodology research (Guerra, 2014; Maranzano, 2002; Martin, 2014) and in commentary (Shuler, 1996; Swanick, 2008; Wise et al., 1984). The findings here support research and commentary in identifying professional growth as a discreet area of concern as it relates to music teacher evaluation. Research new to this area was the examination of music teacher perception regarding the two variables of model and administrator type and their potential interaction. Significant interaction between these two factors as they related to professional growth was encountered in the first question of this domain. Significant differences between model and degree type were also found.

Question 3.1 “You found the evaluation useful towards your professional growth” contained a significant difference for model type in addition to a significant interaction effect between administrator and model type. It is noted that Question 3.2 “You incorporated the suggestions from your evaluation into your future lessons/rehearsals” likewise contained a significant difference for model type with an interaction effect that was marginally significant ($p = .069$).

As was seen in Question 2.6 in Domain 2 concerning an evaluation model best reflecting a respondent's evaluative philosophy, this question likewise resulted in a significant interaction effect. The results for all three of these questions (2.6, 3.1, and 3.2) indicated more favorable perception for Danielson respondents evaluated by someone with a music degree with a downward trend for respondents not evaluated by someone with a music degree, with less favorable perception coming from Stronge respondents evaluated by someone with a music degree with an upward trend for respondents evaluated by someone without a music degree. Or in short, opposite perceptions—the combination of evaluator and model types effecting trends in opposite ways.

Although Question 3.1 received a significant result for model type, this factor should not be examined separately. When an interaction occurs, two or more factors are intertwined (in this case, evaluator and model types), are operating together, and cannot be separated. This statistical principle demonstrates in this instance that although model type significantly influenced teacher perception for Question 3.1, the accompanying interaction effect showed that this was true only for those evaluated by an administrator with a degree in music. Therefore, a question should be posed as it pertains to this interaction: what accounts for the higher perception between respondents of the Danielson model when evaluated by someone with a music degree, and for the lower perception of the Stronge model when evaluated by someone with a music degree, resulting in opposite trends as shown in Figure 6?

Figure 6. Plot for “You found the evaluation useful towards your professional growth”



In an examination of the open-ended responses, there was a greater frequency of discontent expressed from those in Stronge districts in receiving meaningful feedback. This theme emerged among respondents who were evaluated by someone without a music degree—an expected result when compared to previous research. A comparison of the percentage of responses received for this theme between Danielson and Stronge districts may be telling, however: from Danielson respondents, 11 percent of their overall responses were encountered for the theme concerning the absence of meaningful feedback; from Stronge respondents, 29 percent (Figure 11).

If one dismisses this on account of these responses coming only from respondents evaluated in the administrator-without- a-music-degree category, not easily dismissed is the apparent discontent among Stronge respondents when they were evaluated by an administrator with a music degree as well, such discontent occurring at a higher frequency than when compared to the Danielson respondents. For the theme

“Administrator adapts model accordingly” 28 percent of the responses come from Danielson respondents, none for Stronge. For “Administrator incompatibility with discreet music area being assessed”—5 percent for Danielson, 15 percent for Stronge. For “Administrator/Teacher disconnect--administrator not cognizant of current music teacher realities”—11 percent for Danielson, 31 percent for Stronge (Figure 10).

Two possible explanations for these differences harken back to the research of Donald K. Taebel some 24 years earlier that concerned the disparity of results, not of administrator perception, but of the evaluation results of Alabama music teachers when compared to teachers teaching within more classroom specific areas. Taebel, in offering an explanation to the disparate results, posited two: 1) “Music teachers as a group are less competent than other teachers” or 2) the results indicate the evaluation model as defined by the competencies “is not serviceable for evaluating the performance of music teachers” (Taebel, 1990a, p.17). The results between the Danielson and Stronge respondents encountered here could similarly be reduced to 1) Stronge evaluators with a music degree as a group are less competent than equivalent Danielson administrators or 2) the results indicate that Strong administrators may have faced obstacles in contributing to the professional growth to the music teachers they evaluate. Taebel in dismissing the first explanation made a strong case for support of the second. What will account for the differences encountered here?

As was noted in Chapter 2, a hallmark of the Stronge model was its flexibility; aspects of the model could be modified to meet specific school district needs, while on the other hand a trait of the Danielson model was its detailed emphasis on classroom specific elements. In consideration of these differences, it could be thought that opposite

results would have been found, with more favorable teacher perception occurring on account of a model that was more flexible in its use. Yet in the open-ended responses, there was no mention or evidence of this “flexibility” among Stronge respondents. If there was any evidence of flexibility to be seen in the data, it was exhibited among respondents evaluated by the Danielson model.

The absence of expected music-degree-administrator facility in adapting a generic model to better serve music teachers among Stronge respondents when compared to Danielson can be seen as an example of a negative fact. In spite of the flexibility of the Stronge model, Stronge administrators appeared not to adapt the model due to variables that can only be conjectured at this time. What was in evidence however, based on the results obtained from the sample used for this research both quantitatively and qualitatively, was that Danielson respondents on account of some iteration of the model from its original form in the hands of a content specialist indicated a more favorable perception of a model contributing to their professional growth than for Stronge. Returning to Question 2.6, this also may account for the opposite perceptions between Danielson and Stronge respondents in a model best reflecting a respondent’s evaluative philosophy. The perception of a model in this study appears to be influenced by how the model was being used.

Question 3.3 “You would recommend the same administrator in the evaluation of other music teachers for their professional growth” resulted in similar trends for both Danielson and Stronge respondents with significant differences for administrator and model type. The last question, “You would look forward to next evaluation by the same administrator as an opportunity for further professional growth” was significant for

degree type. Again, there is a less favorable perception among the Stronge respondents than when compared to Danielson. Both questions concern administrator type and both contained significant differences for administrator type. Question 3.3 concerning the recommendation of an administrator in the evaluation of other music teachers contained a significant difference for model type as well, with more favorable perception occurring among Danielson respondents. In view of the qualitative findings related to the ANOVA results found here, administrator suitability appeared to be influencing model perception.

Summary of Findings and Discussion

Classroom and rehearsal concerns. The evaluation of classroom and rehearsal specific areas regardless of the two model types used to evaluate them (Danielson or Stronge) or the type of administrator used to evaluate them (administrator with a degree in music or administrator without a degree in music) resulted in no significant differences in attitude and perception between respondents.

Administrator type. Administrator type was identified as a factor in significantly influencing perception as to administrator qualification, contribution to musical development and growth, and the recommendation and anticipation of future evaluations among respondents. The assessment of music content-specific considerations was of greater importance than a model's ability to assess music overall as a curricular area due to the factor of administrator type.

Model type. Model type was identified as a factor in significantly influencing perception as to its usefulness towards professional growth, respondent incorporation of suggestions from an evaluation, and the recommendation of the same administrator in the

evaluation of other music teachers. More favorable perception was found for Danielson respondents than for Stronge, yet with the following caveat:

Interaction. The results indicate that the factors of administrator and model type did not act independently in influencing music teacher attitude and perception regarding evaluative philosophy and professional growth. Rather, they operated together.

Although it appeared that one model was held in greater favor among music teachers over the other, the models' merits appeared to have been influenced by which administrator type was using them, who was adapting them for music teacher evaluation use, and who was not.

Areas for Further Research.

The results suggest that additional research in the domain of classroom competencies for the factors of administrator and evaluator types alone when related directly to music instruction may not be merited.

The presence of significant interaction between the factors of administrator and model type found in this study support that they operated together—not as separate entities operating independently one from the other in the influence of perception. Future research could investigate the degree of influence or gravity each has over the other. Similar to our own earth and moon, though of differing quantities, each influences the other in significant and perceptible ways. Provided a sample could be obtained that was adequate, additional factors could be added, of most interest perhaps being music teacher area of specialization.

If this study had an epicenter for significant results encountered, it fell within the third domain, "Resulting Usefulness and Professional Growth." Containing the least

questions—four as compared to the other domains containing seven each, the domain revealed the majority of the significant results. This raises questions about the importance of professional growth as it pertains to music teacher evaluation when generic models are used, and when individuals of varying ability and resourcefulness employ them.

It remains unclear why Stronge administrators in contrast to Danielson appeared not to adapt their model to better facilitate the evaluation of music teachers. Assuming that there was in fact no difference between the Danielson and Stronge administrators in their effectiveness in contributing toward professional growth, the probability of observing results as great or more would have been less than five percent had the survey been repeated at the time of this study, diminishing to less than one percent for some of the research questions. Provisions were not made in the experimental design for interviewing administrators in any of the evaluative conditions of this study, such provisions having been beyond the study's present scope. Future research could allow for this as it relates to model adaptation and professional growth.

The results demonstrated that administrator/model type interaction as it pertained to the domain of professional growth is an area that merits future investigation owing to the findings of this study. In light of these findings, the creators of the evaluation instruments themselves should take into account that when music administrators are able and given the freedom to adapt their models appropriately for music, then the perceived usefulness of the evaluation process is more favorable than without it.

Bibliography

- Asmus, E. P. & Radocy, R. E. (2006). Quantitative Analysis. In R. Colwell (Ed.), *MENC handbook of research methodologies* (95-175). New York, NY: Oxford.
- Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
- Beaver, M.D. (2002). *Instructional supervision: Perspectives of middle school fine arts teachers*. University of Georgia, ProQuest, UMI Dissertations Publishing, 2002. 0804514.
- Becher, E. A. (2011). *Examining the mismatch between the training and assessment of band and orchestra conductors in Michigan* (Ph.D.). Available from ProQuest Dissertations & Theses A&I. (896957797)
- Bill and Melinda Gates Foundation (2013). Measures of Effective Teaching: 1 - Study Information. ICPSR34771-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2013-09-23.
- Brophy, T. (1993). *Evaluation of music educators: Toward defining an appropriate instrument*. EDRS ED 375 029
- Campbell, D. T. (1975). Assessing the impact of planned social change. In G. Lyons (Ed.), *Social Research and Public Policies: The Dartmouth/OECD Conference*. Hanover, NH: Dartmouth College, The Public Affairs Center, chapter 1, pp 3-45.
- Clark, A. (2015). More NJ teachers get poor reviews in new rating system. *New Jersey Advance Media for NJ.com, June 1, 2015*. Retrieved from http://www.nj.com/education/2015/06/teacher_evaluations.html
- Cochran-Smith, M. (2007). Teacher education: Where are we and where are we going? *Collaborative Action for Change*, ed. Margaret Schmidt. New York: Rowan & Littlefield Education.
- Cody, A. (2011). Alfie Kohn: We have to take back our schools. *Blogs: Education Week, July 27, 2011*. Retrieved from http://blogs.edweek.org/teachers/living-in-dialogue/2011/07/alfie_kohn_we_have_to_take_bac.html
- Collins, I. H. (1996). Assessment and evaluation in music teacher education. *Arts Education Policy Review*, 98(1), 16-21.
- Collins, J. C. (2001). *Good to great: Why some companies make the leap and others don't*. New York, NY: HarperBusiness.

- Colwell, R. (2003). The status of arts assessment: Examples from music. *Arts Education Policy Review*, 105(2), 11-18.
- Colwell, R. (2006). Assessment's potential in music education. In R. Colwell (Ed.), *MENC handbook of research methodologies* (199-269). New York, NY: Oxford.
- Conway, C. M. (2002). Perceptions of beginning teachers, their mentors, and administrators regarding preservice music teacher preparation. *Journal of Research in Music Education*, 50(1), 20-36.
- Cope, P. (2003). OFSTED, fun and learning: A case study of a school music inspection. *British Journal of Music Education*, 20(3), 307-315.
- Council of Chief State School Officers (2011). The InTASC model core teaching standards (April 2011) at a glance. Retrieved from <http://www.ccsso.org/Documents/2011/InTASC%202011%20Standards%20At%20A%20Glance.pdf>
- Council of Chief State School Officers (2013). Interstate Teacher Assessment and Support Consortium InTASC *Model core teaching standards and learning progressions for teachers 1.0: A resource for ongoing teacher development*. Washington, DC.
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Danielson, C., & Association for Supervision and Curriculum Development (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA.
- Danielson, C. (2012). It's your evaluation—collaborating to improve teacher practice. *Education Digest: Essential Readings Condensed for Quick Review*, 77(8), 22-27.
- Danielson, C. (2013a). The Danielson group. *The framework*. Retrieved from <https://danielsongroup.org/framework/>
- Danielson, C. (2013b). General questions about the framework. *The Danielson Group*. Retrieved from <https://danielsongroup.org/questions-about-the-framework-for-teaching/>
- Danielson, C. (2014). Charlotte Danielson's framework for teaching. Retrieved from <http://danielsongroup.org/framework/>
- Darling-Hammond, L., Wise, A. E., Pease, R. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of educational research*, 53(3), 285-328.

- Delaney, D. W. (2007). *Elementary general music teachers' evaluations of instruction: Their own and others'*. Temple University, ProQuest, UMI Dissertations Publishing, 2007. 3255116.
- Doerksen, D. P. (1990). *Guide to evaluating teachers of music performance groups*. Music Educators National Conference.
- Doherty, K. & Jacobs, S. (2013). State of the states 2013. Connect the dots: Using evaluations to teacher effectiveness to inform policy and practice. National Council on Teacher Quality.
- Doyle, K. O., Jr. (1983). *Evaluating teaching*. Lexington, MA: D.C. Heath & Co.
- Duke, R. A. (1987). Observation of applied music instruction: The perceptions of trained and untrained observers. *Collected Work: Applications of Research in Music Behavior*. Pages: 115-124. (AN: 1999-14818).
- Edgar, S. (2012). Communication of expectations between principals and entry-year instrumental music teachers: Implications for music teacher assessment. *Arts Education Policy Review*, 113(4), 136-146.
- Education First (2013). *A primer on common-core aligned assessments*. Retrieved from http://www.education-first.com/files/A_Primer_on_Common_Core-Aligned_Assessments_Education_First.pdf
- Eisner, E. W. (1985). *The educational imagination: On the design and evaluation of school programs* (2nd ed.). New York, NY: Macmillan.
- Elmore, R., Sykes, G., & Spillane, L.P. (1992). Curriculum policy. In P.W. Jackson (Ed.), *Handbook of research on curriculum: A project of the American Education Research Association*. New York: Macmillan.
- Elpus, K. (2011). Merit pay and the music teacher. *Arts Education Policy Review*, 112(4), 180-190. doi:10.1080/10632913.2011.592466
- Finney, J. (2006). The good enough music teacher. *National Association of Music Educators Magazine*, 18, 2-4.
- Fisher, R. (2008). Debating assessment in music education. *Research and Issues in Music Education*, 6(1), 1-10.
- Ferguson, D. (2007). Program evaluations in music education: A review of the literature. *Update: Applications of research in music education*, 25(2) 4-15.

- Gardner, D. P., National Commission on Excellence in Education, & others (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the Secretary of Education*. April 1, 1983. ERIC, Ipswich, MA. Accessed July 28, 2015.
- Geisler, C. D. (2014). *Evaluating music educator effectiveness through student growth: An investigation of the Ohio teacher evaluation system* (Ph.D.). Available from ProQuest Dissertations & Theses A&I. (1560886140).
- Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2010). *Supervision and instructional leadership: A developmental approach* (8th ed.). Boston, MA: Pearson.
- Goddard, H. J. (2004). *Elementary music teachers' and principals' perceptions of music teacher evaluation* (Ph.D.). Available from ProQuest Dissertations & Theses A&I. (305157895).
- Greene, B. (2014). *The fabric of the cosmos: Space, time and the texture of reality*. New York, NY: Vintage Books.
- Guerra, A. (2014). *Perceptions of secondary level ensemble directors regarding teacher evaluations in Texas public schools* (D.M.A.). Available from ProQuest Dissertations & Theses A&I. (1545862255).
- Henninger, J. C. (2002). The effects of knowledge of instructional goals on observations of teaching and learning. *Journal of Research in Music Education*, (1). 37-50.
- Hirokawa, J. O. (2013). *Evaluating music teachers: A comparison of evaluations by observers with varied levels of musical and observational background* (D.M.A.). Available from ProQuest Dissertations & Theses A&I. (1466272559).
- Hodsoll, F. (1998). *Toward civilization: A report on arts education*. Washington, DC: National Endowment for the arts.
- Hourigan, R. M. (2011). Race to the Top: Implications for professional development in arts education. *Arts Education Policy Review*, 112(2), 60-64.
- Iwanicki, E. F. (2001). Focusing teacher evaluations on student learning. *Educational Leadership*, 58(5), 57-59.
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118 (3): 843-77.

- Kertz-Welzel, A. (2008). The implementation of the National Standards for music education in American schools (Jürgen Vogt, Trans.). *Journal of Music Education*, <http://www.zfkm.org/sonder08-kertzwelzel.pdf>.
- 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* (Ed.D.). Available from ProQuest Dissertations & Theses A&I. (276220903).
- Martin, L. D. (2014). *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* (Ph.D.). Available from ProQuest Dissertations & Theses A&I. (1651613076).
- Marzano, R. J., Brandt, R., Hughes, C. S., Jones, B. F., Presseisen, B. Z., Rankin, S. C., & Suhor, C. (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2013). *The Marzano teacher evaluation model*. Retrieved from http://tpep-wa.org/wp-content/uploads/Marzano_Teacher_Evaluation_Model.pdf
- MENC (2010). Assessment in music education position statement. Retrieved 11/2/2010.
- Mooney, J. (2012). School districts comparison shop for teacher evaluation systems. *NJ Spotlight Newsletter*, October 17, 2012. Retrieved from <http://www.njspotlight.com/stories/12/10/17/school-districts-comparison-shop-for-teacher-evaluation-systems/>
- Mooney, J. (2013). Majority of NJ schools opt for widely used teacher-evaluation method. *NJ Spotlight Newsletter*, February 22, 2013. Retrieved from <http://www.njspotlight.com/stories/13/02/21/majority-of-nj-schools-opt-for-widely-used-teacher-evaluation-method/>
- NAfME (2011). NAfME Recommendations for Music Teacher Evaluation. Retrieved from http://smte.us/wp-content/uploads/2011/11/teacher_evaluation_ad_guide.pdf
- NAfME (2012). NAfME teacher evaluation position statement. Retrieved from <http://smte.us/wp-content/uploads/2011/11/NAfME-Teacher-Evaluation-Position-Statement.pdf>
- New Jersey Department of Education (2015a). Frequently asked questions. Retrieved from http://www.state.nj.us/education/genfo/faq/faq_eval.shtml

- New Jersey Department of Education (2015b). New Jersey Department of Education approved teacher practice evaluation instruments as of May 1, 2015, 1-4. Retrieved from <http://www.nj.gov/education/AchieveNJ/teacher/approvedlist.pdf>
- Orzolek, D. C. (2006). The paradox of assessment: Assessment as paradox. *RIME: Research and Issues in Music Education*, 4(1).
- Orzolek, D. C. (2007). Navigating the paradox of assessment in music education. *Collected Work: Assessment in music education: Integrating curriculum, theory, and practice* (pp 37-44). Chicago, IL: GIA Publications.
- Orzolek, D. (2014). Measuring music education: Teacher evaluation in Minnesota. *Music Educators Journal*, 100(4), 22-24. doi:10.1177/0027432114522387
- Overland, C. T. (2014). Teacher evaluation and music education: Joining the national discussion. *Music Educators Journal*, 101(1), 56-62. doi:10.1177/0027432114534448
- Parulekar, M. S. (2014). *Determining criteria for the evaluation of high school band directors: A survey of high school principals and band directors in the state of Ohio* (Ph.D.). Available from ProQuest Dissertations & Theses A&I. (1647259275)
- Perrine, W. M. (2013). Music teacher assessment and Race to the Top: An initiative in Florida. *Music Educators Journal*, 100(1), 39-44. doi:10.1177/0027432113490738
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin Press, Inc.
- Piaget, J. (1928). *The Child's Conception of the World*. London: Routledge and Kegan Paul.
- Pieper, J. (2010). *Leisure, the basis of culture*. San Francisco, CA: Ignatius Press.
- Pierce, J. S. (2005). *Music supervision without a music supervisor in northern New Jersey public schools* (Ed.D.). Available from ProQuest Dissertations & Theses A&I. (305009773).
- Renshaw, J. (2012). Gov. Christie hails signing of tenure reform bill as 'a great day for good teachers.' *The Star-Ledger*, August 7, 2012. Retrieved from http://www.nj.com/news/index.ssf/2012/08/gov_christie_hails_signing_of.html
- Rosenshine, B. (1987). Explicit Teaching and Teacher Training. *Journal of Teacher Education*, 38, (no. 3), 34-36.

- Rudhyar, D. (1982). *The Magic of Tone and the Art of Music*. Boulder, Colorado and London: Shambhala.
- 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, C. P. (1989). Individual differences in perception of applied music teaching feedback. *Psychology of Music*, 17(2), 110-122.
- Shuler, S. C. (1996). Assessing teacher competence in the arts: Should Mr. Holland have gotten the gig? *Arts Education Policy Review*, 98(1), 11.
- Shuler, S. C. (2012). Music education for life: Music assessment, part 2—Instructional improvement and teacher evaluation. *Music Educators Journal*, 98(3), 7-10. doi:10.1177/0027432112439000
- Shulman, P. (2012). State of New Jersey Department of Education. *Educator evaluation system implementation update*. Retrieved from <http://www.nj.gov/education/AchieveNJ/resources/043012Memo.pdf>
- Spradley, J. P. (1980). *Participant observation*. New York, NY: Holt, Rinehart and Winston.
- Stronge, J. H. (2012a). *Stronge teacher effectiveness performance evaluation system*. In association with New Jersey Principals and Supervisors Association. Stronge & Associates Educational Consulting, LLC.
- Stronge, J. H. (2012b). *Stronge teacher effectiveness performance evaluation system: Overview*. Stronge & Associates Educational Consulting, LLC.
- Stronge, J. H. & Tonneson, V. C. (2015). *Stronge+: Teacher effectiveness performance evaluation system handbook*. Stronge & Associates Educational Consulting, LLC.
- Swanwick, K. (2008). The 'good-enough' music teacher. *British Journal of Music Education*, 25(1), 9-22.
- Taebel, D. K. (1990a). *An assessment of the classroom performance of music teachers*. Music Educators National Conference.
- Taebel, D. K. (1990b). *Is evaluation fair to music educators?* Music Educators National Conference.
- Taebel, D. K. (1992). The evaluation of music teachers and teaching. In R. Colwell (Ed.), *Handbook of research on music teaching and learning: A project of the Music Educators National Conference*, (310-329). New York, NY: Schirmer Books.

- U.S. Department of Education (2009). *Race to the Top Program Executive Summary*. Rep. Washington, DC: Web. 28 July 2015.
- Weber, L. (2015, April 2). What do workers want from the boss? *The Wall Street Journal*, At Work. Retrieved from <http://blogs.wsj.com/atwork/2015/04/02/what-do-workers-want-from-the-boss/>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and account for difference in teacher effectiveness. Brooklyn, NY: The New Teacher Project. http://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf
- Wiersma, W. (1988). *The Alabama Career Incentive Program: A statewide effort in teacher evaluation*. Atlanta, GA: Career Incentive Program—Teacher Evaluation Project, Georgia State University.
- Williams, J. (2009). *McREL's teacher evaluation system*. Retrieved from http://www.lakeviewschools.net/uploads/McREL/McREL_TeacherRubricSystem%20FINAL%2007_10.pdf
- Winerip, Michael (2012, January 23). In Race to the Top, the dirty work is left to those on the bottom. *New York Times*. Retrieved from http://www.nytimes.com/2012/01/23/education/in-obamas-race-to-the-top-work-and-expense-lie-with-states.html?_r=0
- Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: The Rand Corporation.
- Xu, X. (2013). Stronge teacher evaluation system: A validation report. Strong & Associates Educational Consulting, LLC. Retrieved from http://www.cesa6.org/effectiveness_project/Validation-Report-of-Stronge-Evaluation-System.pdf

Appendix A

New Jersey Music Teacher Evaluation Survey

The following survey is for the purposes of determining the attitudes of music teachers following an evaluation. Please circle one response for each question.

You, the Teacher

Please indicate the degree to which you believe your last evaluation accurately assessed your own performance within the following areas during your lesson/rehearsal:

	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
Your ability to hold the interest of the class/rehearsal	1	2	3	4	5
Your effectiveness using verbal communication	1	2	3	4	5
Your effectiveness using non-verbal communication	1	2	3	4	5
Your ability to redirect off-task behavior	1	2	3	4	5
Your ability to articulate learning/artistic objectives	1	2	3	4	5
Your ability to motivate students towards learning/artistic objectives	1	2	3	4	5
The consideration of abilities unique to your own teaching/rehearsal style	1	2	3	4	5

Evaluator and Evaluation Instrument

Please indicate the degree to which you agree with the following statements:

	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
The administrator was qualified to evaluate you.	1	2	3	4	5
The person conducting the evaluation makes a meaningful contribution to your musical development and growth.					
Your professional relationship with the administrator is positive.	1	2	3	4	5
The evaluation instrument used to evaluate you was appropriate for music.	1	2	3	4	5
The evaluation instrument used to evaluate you was suited to your area of specialization.	1	2	3	4	5
The evaluation instrument your district uses best reflects your own evaluation philosophy.	1	2	3	4	5
You would welcome this administrator to evaluate you again if you had a choice.	1	2	3	4	5

Resulting Usefulness and Professional Growth

Please indicate the degree to which you agree with the following statements:

	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
You found the evaluation useful towards your professional growth.	1	2	3	4	5
You incorporated the suggestions from your evaluation into your future lessons/rehearsals.	1	2	3	4	5
You would recommend the same administrator in the evaluation of other music teachers for their professional growth.	1	2	3	4	5
You would look forward to the next evaluation by the same administrator as an opportunity for further professional growth.	1	2	3	4	5

1. How many years have you been teaching?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ 21-25
- ☐ 26 and over

2. What grade levels do you teach? Mark all that apply.

- ☐ K
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10
- ☐ 11
- ☐ 12

3. What would best describe your area of specialization?

- ☐ General
- ☐ Vocal
- ☐ Strings
- ☐ Band

4. Is your summative evaluation based on student standardized test score data which is not your own?

- ☐ Yes
- ☐ No

5. Is merit pay tied to your district's evaluation?

- ☐ Yes
- ☐ No

If "yes" to question five, did you receive merit pay for academic year 2014-2015?

- ☐ Yes
- ☐ No

6. What is your school's population?

- ☐ 1-100
- ☐ 101-200
- ☐ 201-300
- ☐ 301-400
- ☐ 401-500
- ☐ 501 and above

7. How many students do you teach weekly?

- ☐ 1-50
- ☐ 51-100
- ☐ 101-200
- ☐ 201 and above

8. Does the administrator who typically evaluates you have a degree in music?

- ☐ Yes
- ☐ No

9. Who typically evaluates you?

- ☐ District music or arts supervisor
- ☐ Principal, assistant principal, or other administrator with a music background
- ☐ Principal, assistant principal, or other administrator with no music background
- ☐ Peer or fellow music teacher

10. What assessment model does your school use?

- ☐ Charlotte Danielson: The Framework for Teaching
- ☐ Classroom Assessment Scoring System
- ☐ Classroom Strategies Scale Model
- ☐ Focal Point Teaching practice Model
- ☐ IMPACT
- ☐ H.E.A.T./Danielson Teacher Evaluation Instrument
- ☐ Insight Core Framework
- ☐ Kenilworth Teacher Evaluation Instrument
- ☐ Lenape Regional Teacher Evaluation Instrument
- ☐ Marzano's Causal Teacher Evaluation Model
- ☐ Mid-continent Research for Education and Learning Teacher Evaluation Standards (McREL)
- ☐ North Star-Academy Teacher Evaluation Rubric
- ☐ Pearson Framework for the Observation of Effective Teaching
- ☐ Rhode Island Model: Teacher Evaluation & Support System
- ☐ Stronge Teacher Effectiveness Performance Evaluation System
- ☐ Teacher Evaluation and Improvement Instrument
- ☐ The College-Ready Promise Teaching Framework
- ☐ The 5D+ Teacher Evaluation Rubric
- ☐ The Marshall Rubrics
- ☐ The Newark Public Schools
- ☐ The New Jersey LoTi Teacher Evaluation
- ☐ The SmartStart TeachElite Evaluation System
- ☐ The Thoughtful Classroom Teacher Effectiveness Framework

11. Were you a member of the committee involved in the decision to select the evaluation model for use in your district?

- ☐ Yes
- ☐ No

12. What is your NJMEA region?

- ☐ North Jersey (Region 1)
- ☐ Central Jersey (Region 2)
- ☐ South Jersey (Region 3)
- ☐ I'm not sure
- ☐ I am not a member of NJMEA

Please supply a written comment about your perception of music teacher evaluation in your district.

Appendix B

Informed Consent Form

Dear Fellow Music Educator,

You as a part of the New Jersey Music Educators Association (NJMEA) membership of approximately 2,000 music educators are being asked to take part in a research study exploring the perceptions and attitudes of music teachers as a result of routine evaluation. This study is being undertaken to better understand the perceived usefulness that the evaluation process has in assisting the professional growth of active music teachers. You are being asked to participate if you are currently employed at least half-time in a K-12 public school setting with a primary job assignment as a music teacher. The survey should take only 5 minutes to complete, and your responses will remain completely anonymous. Your participation in this study is voluntary, and you may withdraw at any time without penalty. The findings of this research are currently planned to be shared on the NJMEA webpage *Tempo Express* by May of 2016. The contact person for this research is NJMEA member Mr. Domecq Smith, a D.M.A. candidate at the Mason Gross School of the Arts, Rutgers University, 732-388-9719. Thank you for your assistance in this research project!

Appendix C

New Jersey Music Teacher Evaluation Survey Hard Copy Introductory Page

New Jersey Music Teacher Evaluation Survey

This study is the first Rutgers approved study investigating music teacher perception of the evaluation process for New Jersey. Be a part of it. Let your voice count. Your responses have indicated that this is an important and neglected area of concern. If you have participated in this research already, T-H-A-N-K Y-O-U ! Please encourage your colleagues to do the same.

There are three ways to access the survey! It is short! –no more than five minutes long!!

1. Complete the survey right here and return to a representative in the lobby! *or*
2. In your email, go to **Tempo Express, All-State Band (02/04/16)** where you will find the Music Teacher Evaluation Survey link to complete online, *or*
3. Go directly to www.surveymonkey.com/r/NJmusicteacherevaluation

Thank you for your participation!



Participant Information and Informed Consent Form

You as a part of the New Jersey Music Educators Association (NJMEA) membership of approximately 2,000 music educators are being asked to take part in a research study exploring the perceptions and attitudes of music teachers as a result of routine evaluation. This study is being undertaken to better understand the perceived usefulness that the evaluation process has in assisting the professional growth of active music teachers in New Jersey. You are being asked to participate if you are currently employed at least half-time in a K-12 public school setting in New Jersey with a primary job assignment as a music teacher. The survey should take only 5 minutes to complete, and your responses will remain completely anonymous. Your participation in this study is voluntary, and you may withdraw at any time without penalty. The findings of this research are currently planned to be shared on the NJMEA webpage *Tempo Express* by May of 2016. The contact person for this research is NJMEA member Mr. Domecq Smith, a D.M.A. candidate at the Mason Gross School of the Arts, Rutgers University, 732-388-9719. Thank you for your assistance in this research project!

Appendix D

Distribution T-shirts, Front and Back, for NJMEA February 2016 Convention



Appendix E

New Jersey Counties and Public School Districts Determined to Utilize the Stronge Teacher Effectiveness Performance Evaluation System for Phase 3 of Survey Dissemination

Bergen County

Cliffside Park	Edgewater	Fair Lawn
Glen Rock	Hillsdale	Montvale
Oradell	Palisades Park	Paramus
Ramapo	Indian Hills	Ramsey
Ridgefield	Ridgefield Park	Ridgewood
Riverdell	River Edge	River Vale
Rochelle Park	Saddle River	Woodcliff Lake

Burlington County

Burlington	Lumberton	Mount Holly
Rancocas Valley	Shamong	Westhampton

Camden County

Black Horse Pike	Haddon	Oaklyn
------------------	--------	--------

Essex County

Cedar Grove	Essex Fells	Essex Regional ESC
Fairfield	Link Community Charter	North Caldwell
Pride Academy Charter	Roseland	

Hudson County

East Newark	Great Futures Charter HS
-------------	--------------------------

Hunterdon County

Alexandria	North-Hunterdon-Voorhees
------------	--------------------------

Mercer County

East Windsor

Middlesex County

Dunellen
Middlesex Regional ESC
South River

Metuchen
South Amboy
Spotswood

Metuchen Christian
South Brunswick

Monmouth County

Belmar

Bradley Beach

Roosevelt

Morris County

Dover
Netcong

Morris Hills
Randolph

Mount Olive
West Morris

Passaic County

Hawthorne
North Haledon
Ringwood Christian

Holland Charter
Passaic County ESC
Wayne

Lakeland
Ringwood

Somerset County

Bound Brook
Somerville

Branchburg

Hillsborough

Sussex County

Hopatcong

Sussex County ESC

Vernon

Union County

Rahway

Union

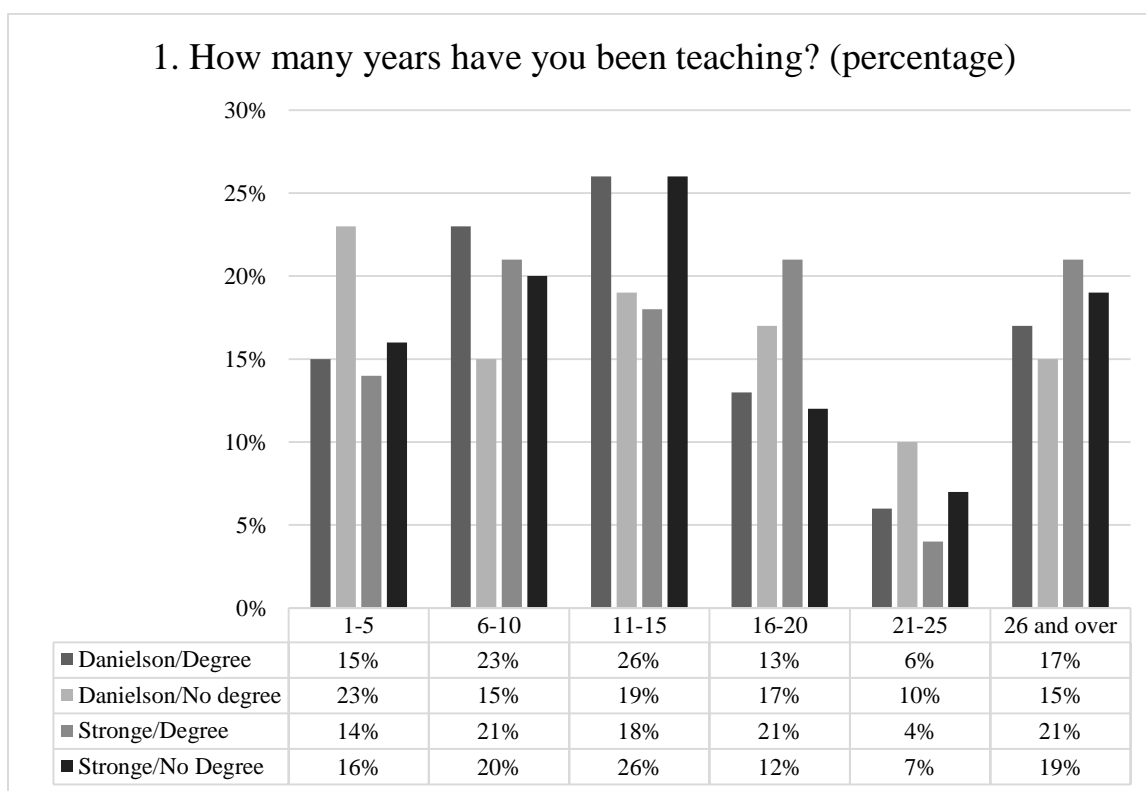
Appendix F

Demographic Data: Responses by Frequency and Percentage

1. How many years have you been teaching? (frequency)

	1-5	6-10	11-15	16-20	21-25	26 and over	Total
Danielson/Degree	7	11	12	6	3	8	47
Danielson/No Degree	34	21	28	25	15	22	145
Stronge/Degree	4	6	5	6	1	6	28
Stronge/No Degree	13	17	22	10	6	16	84
Total	58	55	67	47	25	52	304

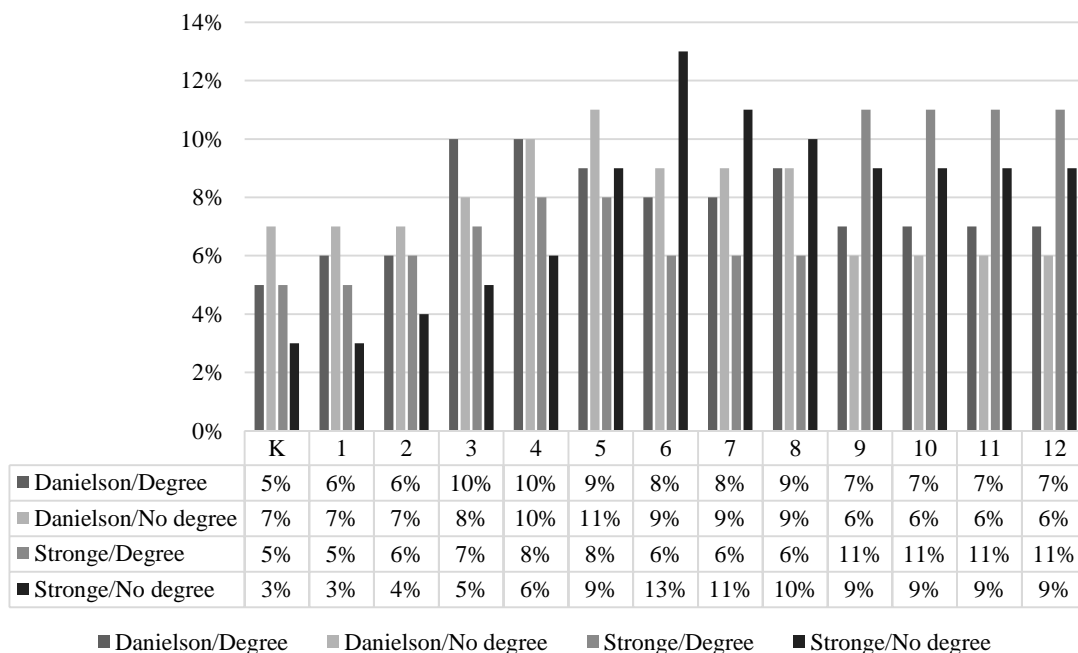
1. How many years have you been teaching? (percentage)



2. What grade levels do you teach? Mark all that apply. (frequency)

	K	1	2	3	4	5	6	7	8	9	10	11	12	Total
Danielson/Degree	11	12	13	22	22	20	16	18	19	15	15	15	15	213
Danielson/No Degree	50	51	52	55	71	76	67	63	64	41	40	40	40	710
Stronge/Degree	6	6	8	9	10	10	7	8	7	14	14	14	14	127
Stronge/No Degree	13	13	17	19	25	35	53	43	42	36	36	36	36	404
Total	80	82	90	105	128	141	143	132	132	106	105	105	105	1454

2. What grade levels do you teach? Mark all that apply. (percentage)

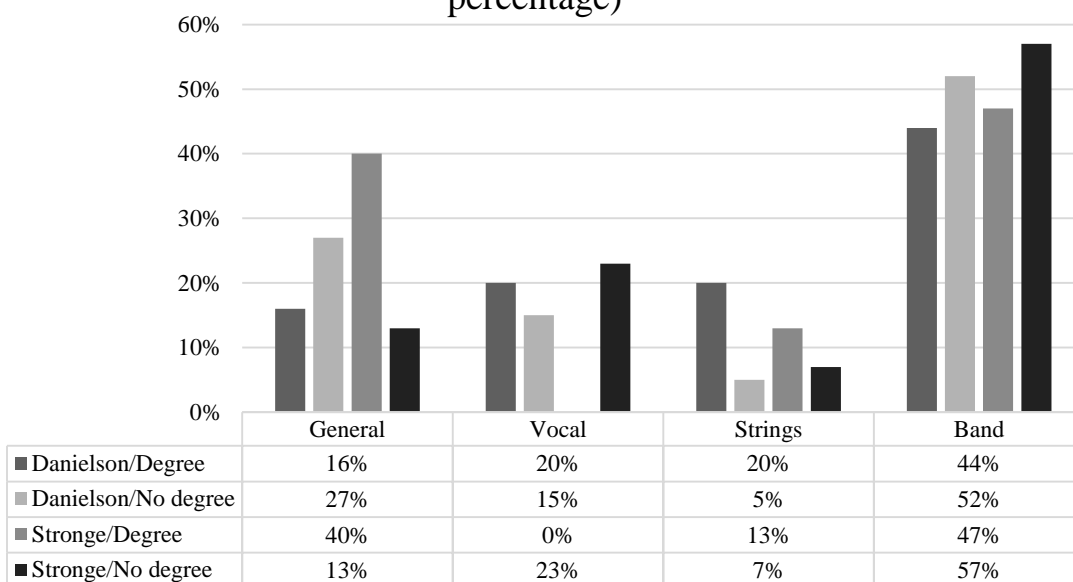


3. What is your area of specialization? (online responses by frequency)

	General	Vocal	Strings	Band	Total
Danielson/Degree	4	5	5	11	25
Danielson/No Degree	25	14	5	48	92
Stronge/Degree	6	0	2	7	15
Stronge/No Degree	4	7	2	17	30
Total	39	26	14	83	162

Note. Data for Question 3 “What is your area of specialization” is listed as two categories: “Online Responses” and “Paper Responses.” This is due to the online version of the question having allowed for one response only, whereas respondents using the paper version of the survey indicated multiple areas of specialization.

3. What is your area of specialization? (online responses by percentage)



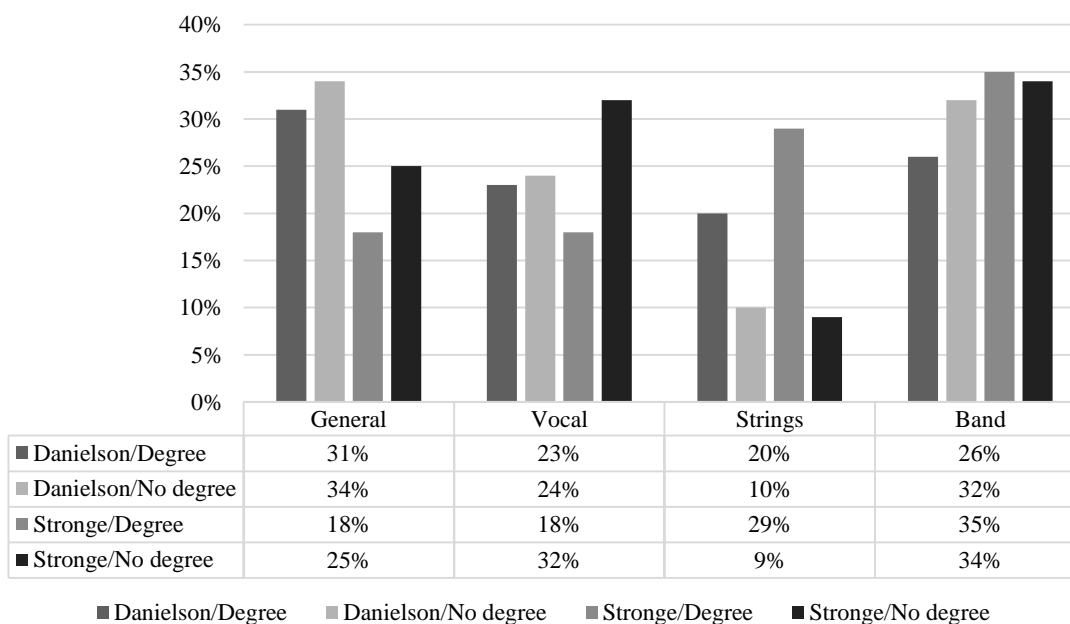
■ Danielson/Degree ■ Danielson/No degree ■ Stronge/Degree ■ Stronge/No degree

3. What is your area of specialization? (paper responses by frequency)

	General	Vocal	Strings	Band	Total
Danielson/Degree	12	9	8	10	39
Danielson/No Degree	28	20	8	27	83
Stronge/Degree	3	3	5	6	17
Stronge/No Degree	22	28	7	30	87
Total	65	60	28	73	226

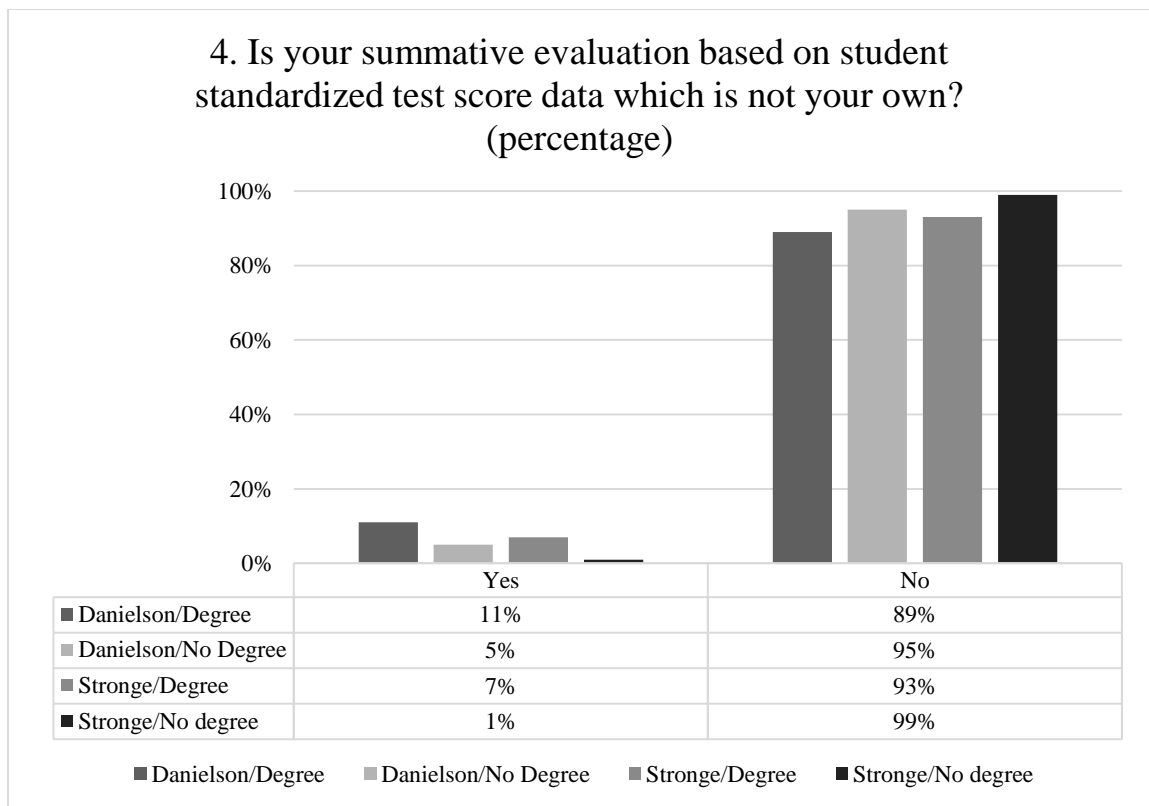
Note. Data for Question 3 “What is your area of specialization” is listed as two categories: “Online Responses” and “Paper Responses.” This is due to the online version of the question having allowed for one response only, whereas respondents using the paper version of the survey indicated multiple areas of specialization.

3. What is your area of specialization? (paper responses by percentage)



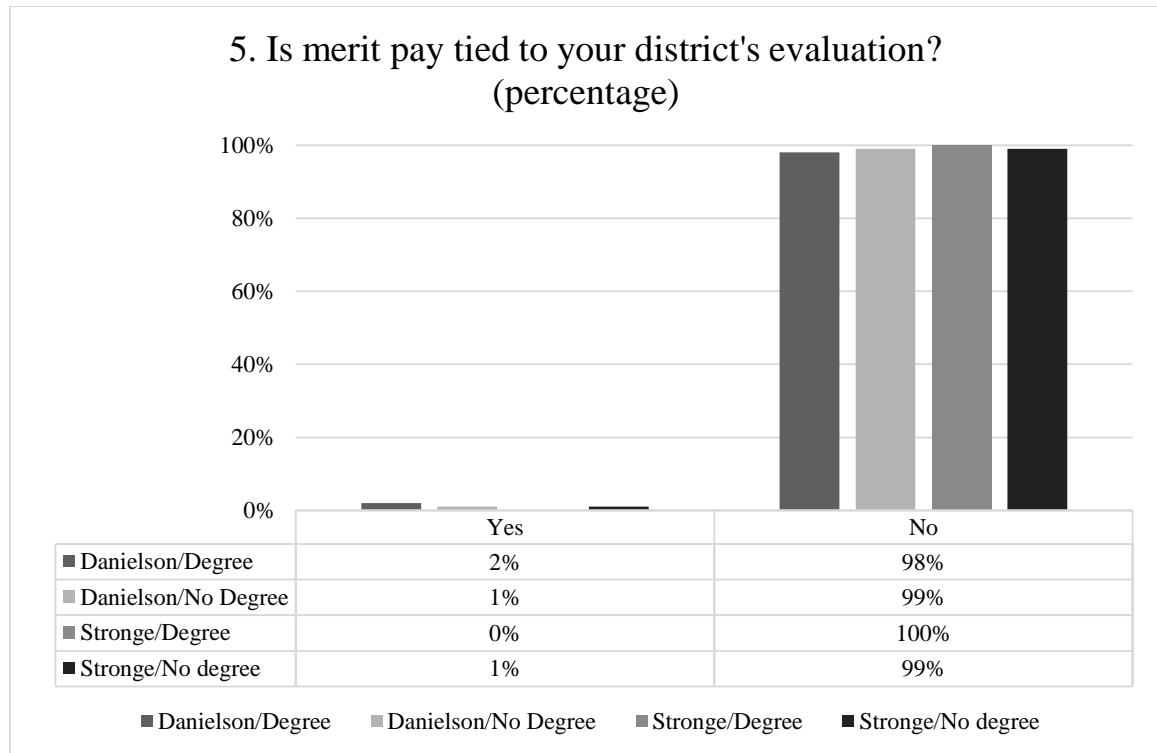
4. Is your summative evaluation based on student standardized test score data which is not your own? (frequency)

	Yes	No	Total
Danielson/Degree	5	42	47
Danielson/No Degree	7	138	145
Stronge/Degree	2	26	28
Stronge/No Degree	1	86	87
Total	15	292	307



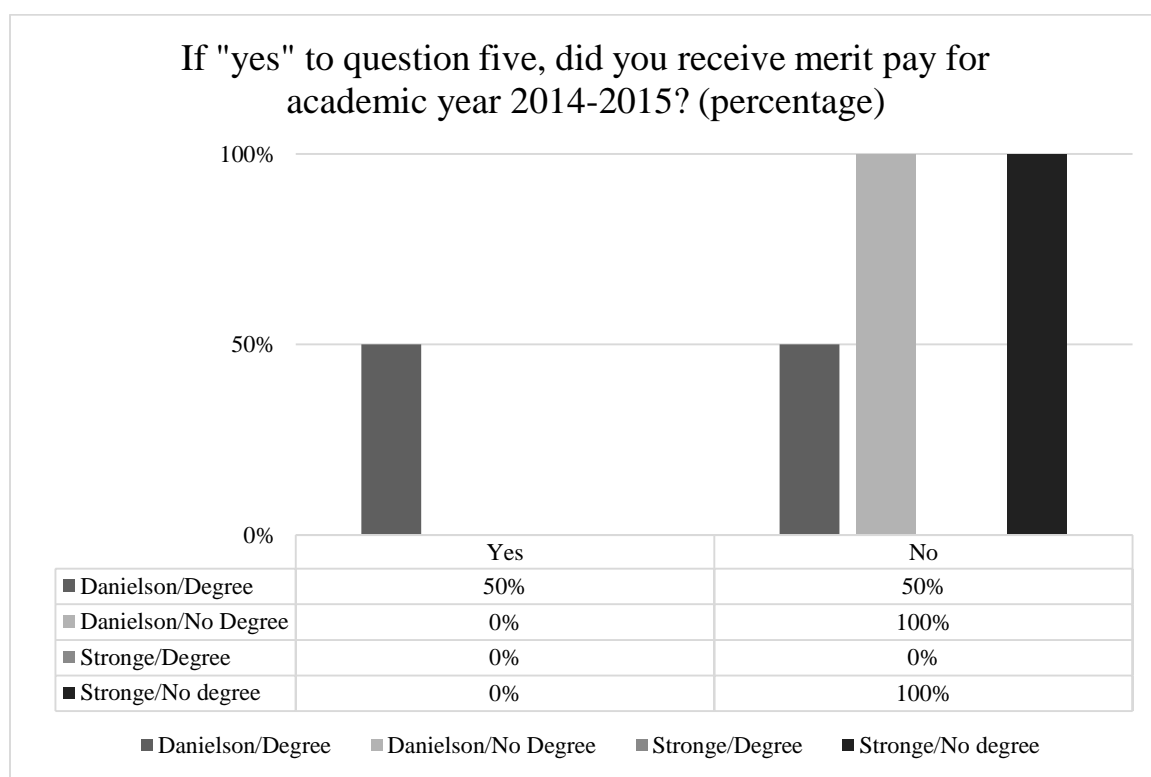
5. Is merit pay tied to your district's evaluation? (frequency)

	Yes	No	Total
Danielson/Degree	1	45	46
Danielson/No Degree	1	143	144
Stronge/Degree	0	28	28
Stronge/No Degree	1	86	87
Total	3	302	305



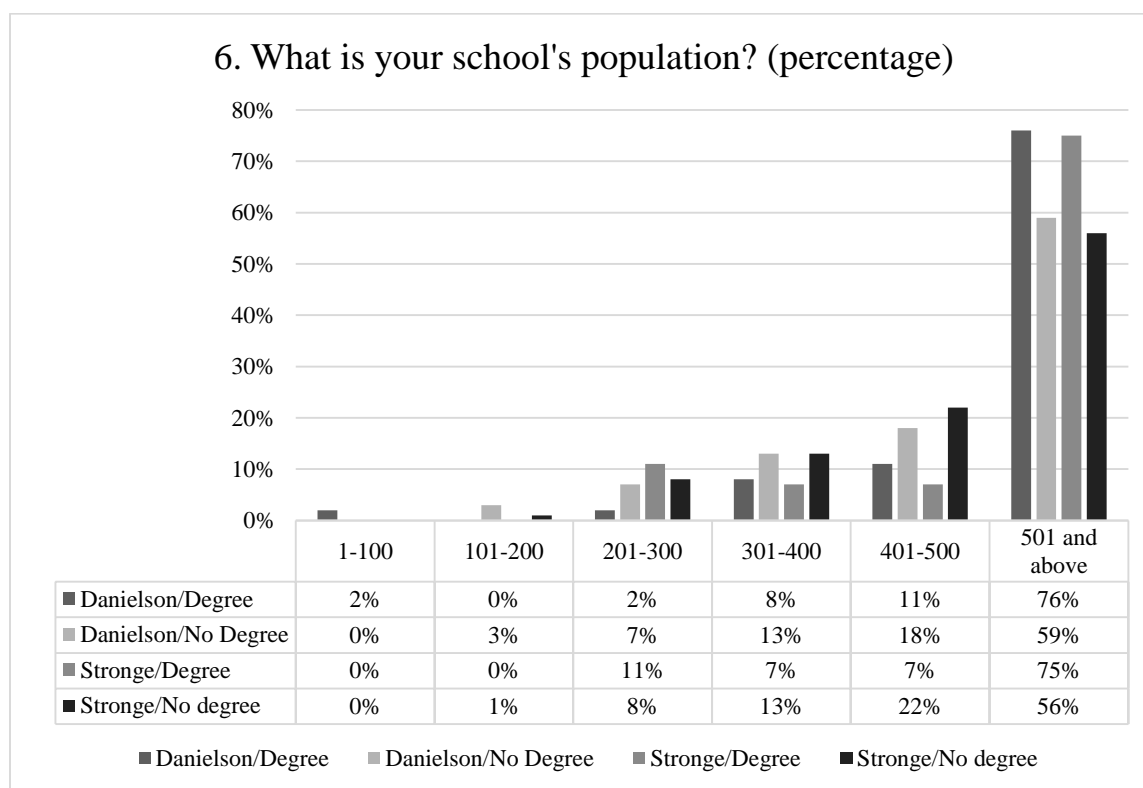
5. If “yes” to question 5, did you receive merit pay for academic year 2014-2015? (frequency)

	Yes	No	Total
Danielson/Degree	1	1	2
Danielson/No Degree	0	1	1
Stronge/Degree	0	0	0
Stronge/No Degree	0	1	1
Total	1	3	4



6. What is your school's population? (frequency)

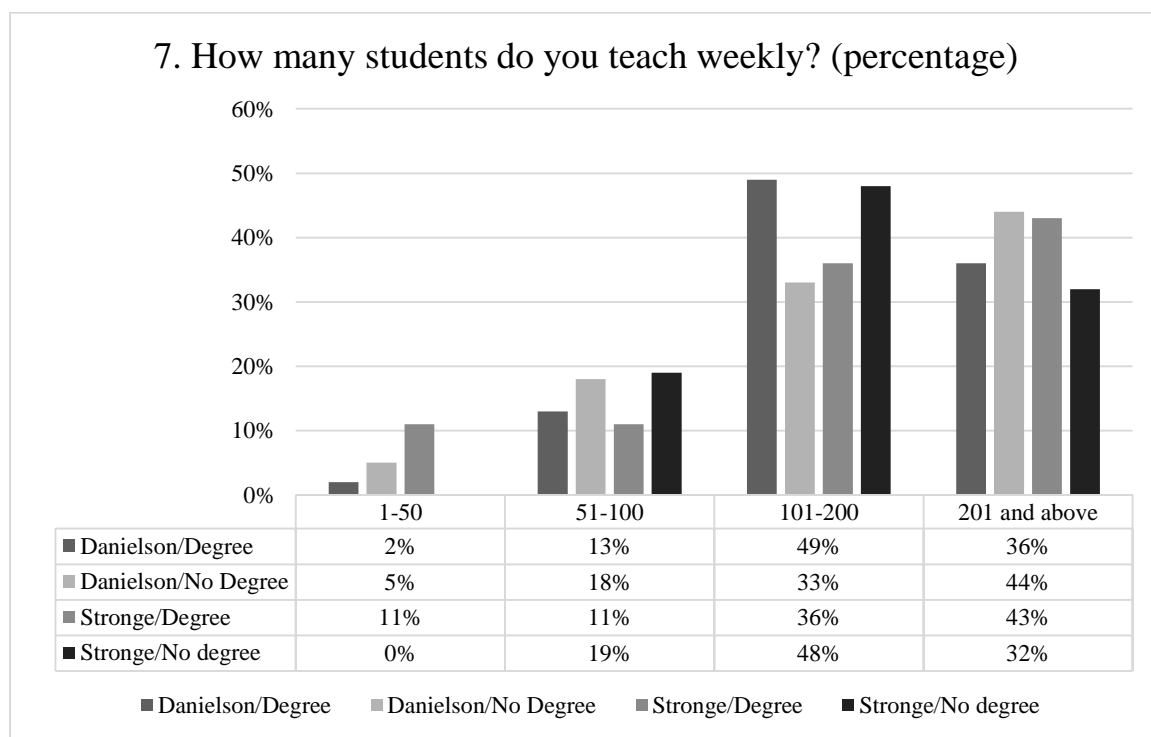
	1-100	101-200	201-300	301-400	401-500	501 and above	Total
Danielson/Degree	1	0	1	4	5	36	47
Danielson/No Degree	0	4	10	19	26	87	146
Stronge/Degree	0	0	3	2	2	21	28
Stronge/No Degree	0	1	7	11	19	49	87
Total	1	5	21	36	52	193	308



7. How many students do you teach weekly? (frequency)

	1-50	51-100	101-200	201 and above	Total
Danielson/Degree	1	6	23	17	47
Danielson/No Degree	7	27	48	64	146
Stronge/Degree	3	3	10	12	28
Stronge/No Degree	0	17	42	28	87
Total	11	53	123	121	308

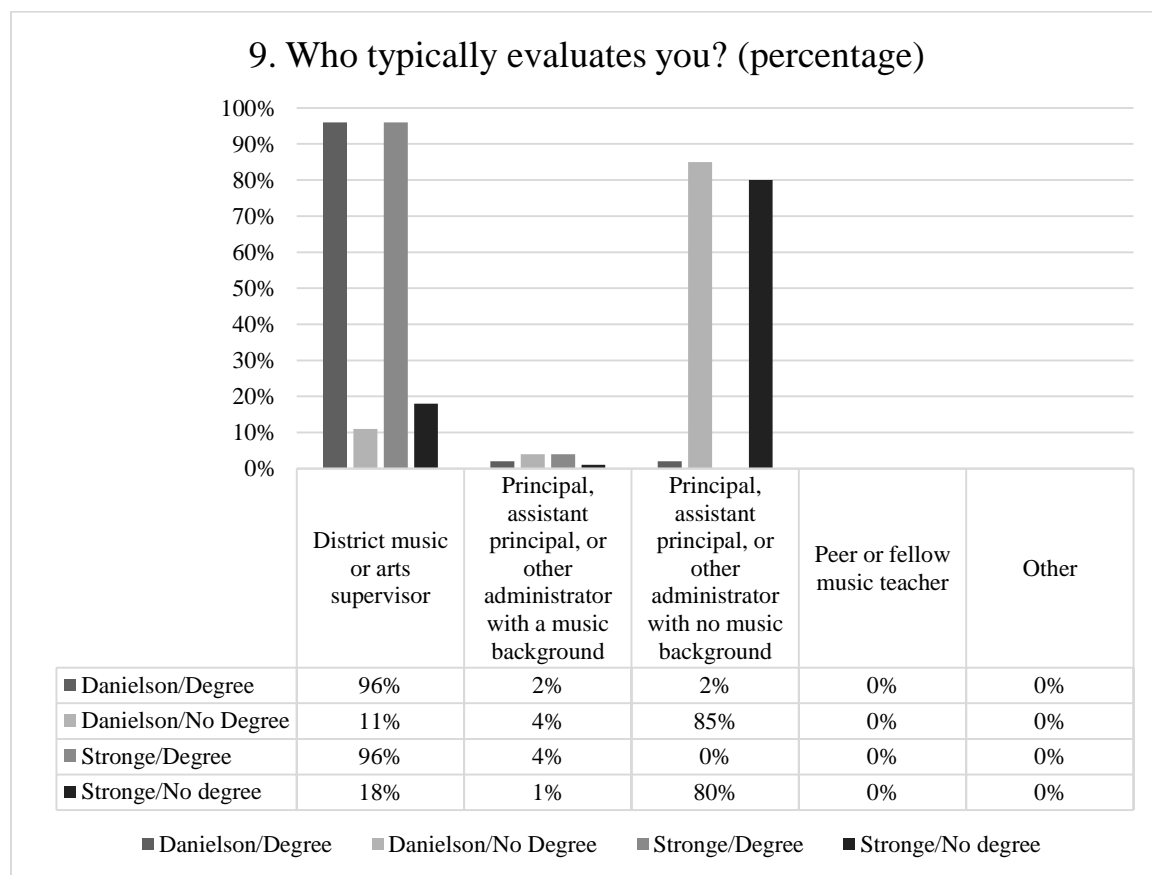
7. How many students do you teach weekly? (percentage)



9. Who typically evaluates you? (frequency)

	District music or arts supervisor	Principal, assistant principal, or other administrator with a music background	Principal, assistant principal, or other administrator with no music background	Peer or fellow music teacher	Other	Total
Danielson/Degree	45	1	2	—	—	47
Danielson/No Degree	16	6	85	—	—	146
Stronge/Degree	27	1	—	—	—	28
Stronge/No Degree	16	1	70	—	—	87
Total	104	9	195	—	—	308

Note. A dash indicates no response.



10. What assessment model does your school use? (frequency)

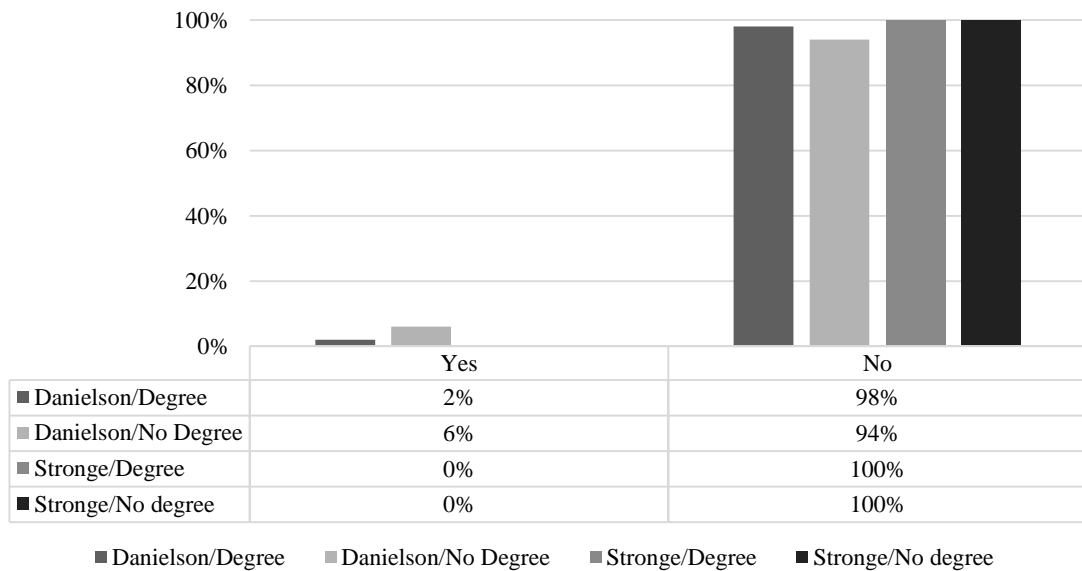
	Administrator with music degree	Administrator without music degree	Total
Classroom Strategies Scale Model	1	1	2
Classroom Assessment Scoring System	—	—	—
Focal Point Teaching Practice Model	—	—	—
IMPACT	—	—	—
H.E.A.T./Danielson Teacher Evaluation Instrument	—	2	2
Insight Core Framework	—	—	—
Kenilworth Teacher Evaluation Instrument	—	—	—
Lenape Regional Teacher Evaluation Instrument	—	—	—
Marzano Causal Teacher Evaluation Model	4	23	27
Mid-continent Research for Education and Learning Teacher Evaluation Standards (McREL)	7	12	19
North Star-Academy Teacher Evaluation Rubric	—	—	—
Pearson Framework for the Observation of Effective Teaching	—	1	1
Rhode Island Model	—	—	—
Teacher Evaluation and Improvement Instrument	—	3	3
The College-Ready Promise Teaching Framework	—	—	—
The 5D+ Teacher Evaluation Rubric	—	—	—
The Marshall Rubrics	7	11	18
The Newark Public Schools	—	5	5
The New Jersey LoTi Teacher Evaluation	—	—	—
The SmartStart TeachElite Evaluation System	—	—	—
The Thoughtful Classroom Teacher Effectiveness Framework	—	3	3
Total	19	61	80

Note. This table reports non Danielson FfT and Stronge TEPES responses only.

11. Were you a member of the committee responsible for selecting the evaluation model used in your district? (frequency)

	Yes	No	Total
Danielson/Degree	1	46	47
Danielson/No Degree	9	135	144
Stronge/Degree	0	28	28
Stronge/No Degree	0	87	87
Total	10	296	306

11. Were you a member of the committee responsible for selecting the evaluation model used in your district? (percentage)

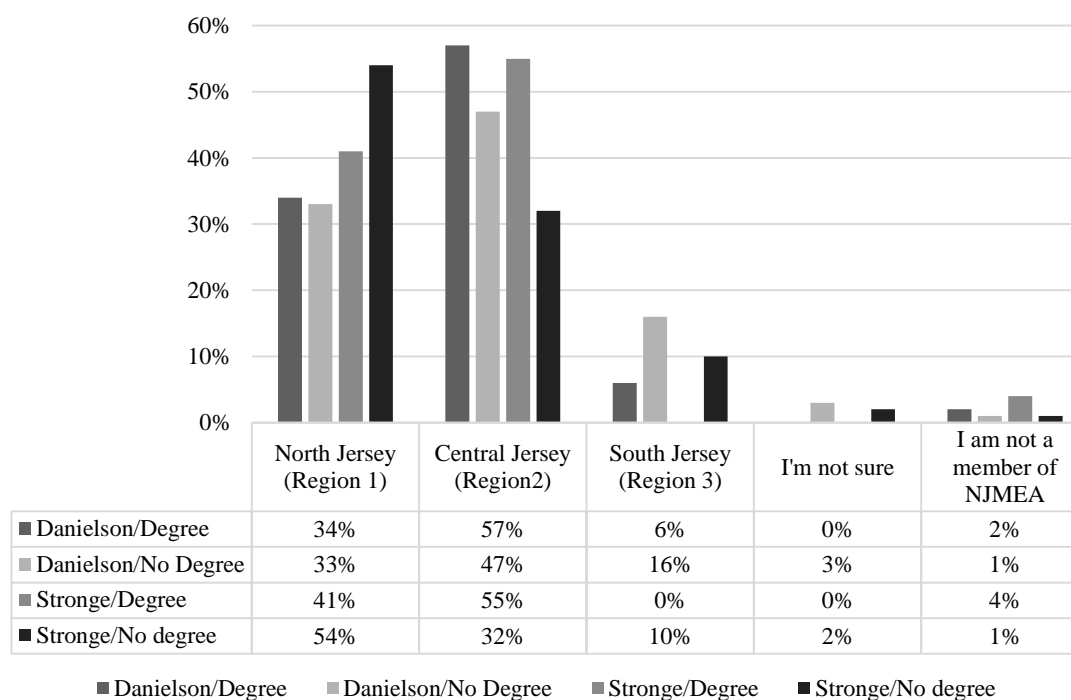


11. What is your NJMEA region? (frequency)

	North Jersey (Region 1)	Central Jersey (Region 2)	South Jersey (Region 3)	I'm not sure	I am not a member of NJMEA	Total
Danielson/Degree	16	27	3	—	1	47
Danielson/No Degree	48	69	24	4	1	146
Stronge/Degree	11	15	—	—	1	27
Stronge/No Degree	46	27	9	2	1	85
Total	121	138	36	6	4	305

Note. A dash indicates no response.

12. What is your NJMEA region? (percentage)



			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
1.1	Your ability to hold the interest of the class/rehearsal	Danielson	47	4.09	1.07	146	4.04	0.88	193	4.05	0.93
		Stronge	28	4.11	1.14	87	4.23	0.94	115	4.20	1.00
			75	4.09	1.10	233	4.11	0.91	308	4.11	0.96
			Degree $p = .77$ Model $p = .43$ Degree*Model $p = .53$								
1.2	Your effectiveness using verbal communication	Danielson	47	4.04	1.07	146	4.08	0.75	193	4.07	0.84
		Stronge	28	4.18	1.00	87	4.24	0.88	115	4.23	0.91
			75	4.09	1.05	233	4.14	0.81	308	4.13	0.87
			Degree $p = .69$ Model $p = .21$ Degree*Model $p = .90$								
1.3	Your effectiveness using non-verbal communication	Danielson	47	3.81	1.08	146	3.72	0.99	193	3.74	1.23
		Stronge	28	3.79	1.18	87	3.68	1.24	115	3.70	1.22
			75	3.80	1.12	233	3.70	1.09	308	3.73	1.10
			Degree $p = .52$ Model $p = .83$ Degree*Model $p = .95$								
1.4	Your ability to redirect off-task behavior	Danielson	47	4	0.99	146	3.85	0.89	193	3.89	0.93
		Stronge	28	3.71	1.19	87	3.94	1.11	115	3.89	1.13
			75	3.89	1.09	233	3.88	0.98	308	3.89	1.01
			Degree $p = .78$ Model $p = .49$ Degree*Model $p = .17$								

			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
2.5	The evaluation instrument used to evaluate you was suited to you area of specialization	Danielson	47	2.77	1.19	145	2.29	1.01	192	2.41	1.08
		Stronge	27	2.74	1.26	87	3.51	0.96	114	2.56	1.04
			74	2.76	1.22	232	2.37	1.00	306	2.46	1.07

Degree $p = .02$ Model $p = .51$ Degree*Model $p = .41$

			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
2.6	The evaluation instrument your district uses best reflects your own evaluation philosophy	Danielson	47	2.74	1.12	146	2.40	1.01	193	2.48	1.05
		Stronge	27	2.37	1.13	87	2.60	0.93	114	2.54	0.98
			74	2.61	1.14	233	2.47	0.99	307	2.50	1.03

Degree $p = .67$ Model $p = .54$ Degree*Model $p = .04$

			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
2.7	You would welcome the administrator to evaluate you again if you had a choice	Danielson	47	3.89	1.15	146	3.56	1.14	193	3.64	1.05
		Stronge	27	4.11	1.07	87	3.79	1.04	114	3.87	1.06
			74	3.97	1.13	233	3.65	1.11	307	3.73	1.12

Degree $p = .04$ Model $p = .15$ Degree*Model $p = .96$

Appendix I

Summary Statistics for Domain 3, “Resulting Usefulness and Growth.”

			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
3.1	You found the evaluation useful towards your professional growth	Danielson	47	3.17	1.12	146	2.53	1.06	193	2.68	1.11
		Stronge	28	2.43	1.18	87	2.61	1.03	115	2.57	1.07
			75	2.89	1.20	233	2.56	1.05	308	2.64	1.10

Degree $p = .12$ Model $p = .03$ Degree*Model $p = .006$

			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
3.2	You incorporated the suggestions from your evaluation into your future lessons/rehearsals	Danielson	47	3.47	1.01	146	2.95	1.09	193	3.07	1.09
		Stronge	28	2.89	1.29	87	2.92	1.06	115	2.91	1.12
			75	3.25	1.16	233	2.94	1.08	308	3.01	1.11

Degree $p = .10$ Model $p = .047$ Degree*Model $p = .07$

			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
3.3	You would recommend the same administrator in the evaluation of other music	Danielson	47	3.62	1.10	146	2.60	1.15	193	2.84	1.22
		Stronge	28	2.96	1.52	87	2.51	1.08	115	2.62	1.22
			75	3.37	1.31	233	2.56	1.13	308	2.76	1.23

Degree $p = .001$ Model $p = .02$ Degree*Model $p = .08$

			Degree			No Degree					
			<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
3.4	You would look forward to the next evaluation by the same administrator as an opportunity	Danielson	47	3.40	1.21	146	2.71	1.19	193	2.88	1.23
		Stronge	28	2.89	1.23	87	2.72	1.04	115	2.77	1.09
			75	3.21	1.25	233	2.72	1.14	308	2.84	1.18

Degree $p = .008$ Model $p = .12$ Degree*Model $p = .10$

Appendix J

RUTGERS

Office of Research and Regulatory Affairs
Arts and Sciences IRB
Rutgers, The State University of New Jersey
335 George Street / Liberty Plaza / Suite 3200
New Brunswick, NJ 08901

orra.rutgers.edu/artsci

732-235-9806

December 21, 2015

P.I. Name: Smith
Protocol #: E16-239

Domecq Smith
Mason Gross School of the Arts
33 Livingston Avenue
New Brunswick NJ 08901

Dear Domecq Smith:

This project identified below has been approved for exemption under one of the six categories noted in 45 CFR 46, and as noted below:

Protocol Title: "Music Teacher Evaluation: New Jersey Music Teachers Perceptions of Evaluation Models and Evaluator Types"

Amendment Exemption Date: 12/21/2015 **Exempt Category:** 2

This exemption is based on the following assumptions:

- **This Approval** - The research will be conducted according to the most recent version of the protocol that was submitted.
- **Reporting** - ORSP must be immediately informed of any injuries to subjects that occur and/or problems that arise, in the course of your research;
- **Modifications** - Any proposed changes MUST be submitted to the IRB as an amendment for review and approval prior to implementation;
- **Consent Form (s)** - Each person who signs a consent document will be given a copy of that document, if you are using such documents in your research. The Principal Investigator must retain all signed documents for at least three years after the conclusion of the research;

Additional Notes:	<ul style="list-style-type: none"> • Amendment to Exemption Granted on 12/21/2015 for Change in Title from "Music Teacher Evaluation: Administrator and Evaluation Model Perceptions of New Jersey Music Teachers" to "Music Teacher Evaluation: New Jersey Music Teachers Perceptions of Evaluation Models and Evaluator Types"; Revised Survey to Alter Some Items & for Clarity Per Dissertation Committee
--------------------------	--

Failure to comply with these conditions will result in withdrawal of this approval.
Please note that the IRB has the authority to observe, or have a third party observe, the consent process or the research itself.
The Federal-wide Assurance (FWA) number for the Rutgers University IRB is FWA00003913; this number may be requested on funding applications or by collaborators.

Sincerely yours,



Acting For--
Beverly Teppe, Ph.D.
Professor, Department of Food Science
IRB Chair, Arts and Sciences Institutional Review Board
Rutgers, The State University of New Jersey

cc: Dr. William Berz, Mason Gross School of the Arts

(MW)