

2015

A Correlational Study of Teacher Effectiveness: Evaluation Instrument and Value-Added Model

Danielle Simmons Banister
Gardner-Webb University

Follow this and additional works at: https://digitalcommons.gardner-webb.edu/education_etd



Part of the [Education Commons](#)

Recommended Citation

Banister, Danielle Simmons, "A Correlational Study of Teacher Effectiveness: Evaluation Instrument and Value-Added Model" (2015). *Education Dissertations and Projects*. 121.
https://digitalcommons.gardner-webb.edu/education_etd/121

This Dissertation is brought to you for free and open access by the School of Education at Digital Commons @ Gardner-Webb University. It has been accepted for inclusion in Education Dissertations and Projects by an authorized administrator of Digital Commons @ Gardner-Webb University. For more information, please see [Copyright and Publishing Info](#).

A Correlational Study of Teacher Effectiveness: Evaluation Instrument and Value-Added
Model

By
Danielle Simmons Banister

A Dissertation Submitted to the
Gardner-Webb School of Education
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

Gardner-Webb University
2015

Approval Page

This dissertation was submitted by Danielle Simmons Banister under the direction of the persons listed below. It was submitted to the Gardner-Webb University School of Education and approved in partial fulfillment of the requirements for the degree of Doctor of Education at Gardner-Webb University.

Kelly Taylor Clark, Ed.D.
Committee Chair

Date

Sydney Brown, Ph.D.
Committee Member

Date

David Shellman, Ed.D.
Committee Member

Date

Jeffrey Rogers, Ph.D.
Dean of the Gayle Bolt Price School
of Graduate Studies

Date

Dedication

To you, sweet boy.

You will always be my biggest accomplishment.

My greatest source of pride.

I love you to the moon and back, Bradlee!

Acknowledgements

I wish to extend my appreciation to the teachers and staff of the school system for your dedication and commitment to educating the 144,000 children who come to school each and every day. I am so grateful to work in the same district in which I was educated that truly puts students above all else. Thank you to the district, especially Dr. Kris Zorigian and the Talent Management team, for the assistance and support you provided throughout this process.

My sincerest thanks are also extended to my Educational Leadership cohort at the Gardner-Webb University Charlotte campus. I cannot express how much I have appreciated your perspective, wisdom, friendship, and inspiration over the last 4 years.

I would also like to thank Dr. Kelly Taylor Clark, my dissertation chair, for your support, partnership, and expertise throughout this process. Additionally, thank you to my committee members, Dr. Sydney Brown and Dr. David Shellman; my professors, Dr. Kelly Gwaltney, Dr. John Reynolds, Dr. Dennis Triplett, Dr. Steve Bingham; and also to Deans Doug Eury and Jeff Rogers.

Words cannot truly articulate my gratitude for my family, but I owe my most heartfelt love and thanks to my husband, Brandon Banister; son, Bradlee Banister; stepchildren, Madison and Bryson Banister; parents, Brad and Renee Simmons; and brother, Darrall Simmons. Thank you for your support, accountability, and patience during those early mornings, late nights, absent weekends, and missed events. I appreciate your unselfish encouragement as I completed this doctoral program and my dissertation. I have only ever wanted to make you proud of me. I love you!

Thank you all from the very bottom of my heart.

Abstract

A Correlational Study of Teacher Effectiveness: Evaluation Instrument and Value-Added Model. Banister, Danielle Simmons, 2015: Dissertation, Gardner-Webb University, Teacher Effectiveness/Teacher Evaluation/Principal Observation/Value-Added Model

This correlational study was conducted to determine the relationship between two measures of teacher effectiveness in a southeastern state in the United States. The state utilizes a teacher evaluation instrument that rates teachers based on principal observations on five standards. Additionally, a sixth standard is populated with data from a value-added model that measures a teacher's impact on student learning based on student achievement on standardized tests. This study aimed to compare methods used to assess quality teaching. As the teacher has the greatest impact on student achievement, educational agencies and districts have focused efforts on improving teacher performance. However, there currently is not a single instrument that stakeholders agree would quickly and accurately assess teacher effectiveness, necessitating the investigation of evaluation systems and processes for identifying effective teaching.

Research was collected in a large, urban school district in the state to determine the relationship between the two measures within the context of a single school district. The value-added data and the state teacher evaluation instrument data were analyzed among the teachers of tested subjects in Grades 4-12 in the school district to determine if there was a correlation between the ratings provided by each of the measures. Spearman's rank-order coefficient was used to analyze the relationship between scores. The results demonstrated were negligible to weak correlations between the teacher evaluation instrument standards and EVAAS scores. Limitations, recommendations, and implications for future research were included with the findings.

Table of Contents

Chapter 1: Introduction	1
Statement of the Problem.....	2
Conceptual Framework for the Study.....	2
Purpose of the Study	5
Research Questions.....	5
Definition of Terms.....	6
Summary	7
Chapter 2: Literature Review	9
Introduction.....	9
History of Teacher Evaluation	10
Recent Changes to Teacher Evaluation	17
State Changes in Evaluation	25
Recent District Initiatives	36
Summary	39
Chapter 3: Methodology	41
Introduction.....	41
Problem.....	41
Purpose.....	42
Population	42
Collection and Analysis of Data	45
Limitations	51
Delimitations.....	53
Summary	54
Chapter 4: Results	55
Introduction	55
Research Questions	55
Data Collection Process	56
Sample Demographics	57
Monotonic Relationship	59
Findings	60
Additional Correlations	64
Summary	75
Chapter 5: Discussion	76
Introduction	76
Summary of Findings	76
Research Questions	78
Recommendations	87
Implications for Future Research	89
Limitations	93
Summary	94
References	96
Appendices	
A State Teacher Evaluation Rubric	103
B Data Scatterplots	115

Tables	
1	Breakdown of Student Population43
2	Student Demographics of District44
3	Teacher Demographics of District45
4	Evaluation Plan51
5	Sample Teacher Demographics58
6	Teacher Sample Size59
7	Determining the Strength of the Relationship60
8	Spearman’s Coefficient64
9	School-Level Sample Sizes65
10	School-Level Correlations67
11	Gender Sample Sizes68
12	Gender Correlations69
13	Ethnicity Sample Sizes70
14	Ethnicity Correlations73
15	Career Status Sample Sizes.....74
16	Career Status Correlations75
Figure	
	Monotonic and Non-Monotonic Relationships.....49

Chapter 1: Introduction

Study after study being conducted in the United States suggests that there is a crisis in our education system, as students are not achieving. According to the 2011 National Assessment of Educational Progress (NAEP) study, students in the United States are not learning the skills and knowledge they need to succeed in the 21st century (National Center for Education Statistics, 2012). Approximately two-thirds of eighth-grade students cannot currently read proficiently, and the same amount of students did not meet proficiency on recent course exams. According to recent studies measuring fourth- and eighth-grade academic achievement in math and science, there is a trend among students in the United States to perform worse as they age (National Center for Education Statistics, 2012).

Annually, 1.1 million students drop out of school, bringing the national dropout rate average to 27%. For African-American and Hispanic students, dropout rates are near 40%. Of those students who do graduate, many are not prepared for college as only one in four students graduates ready for college in English, reading, math, and science (ACT, 2011). This trend continues as less than half of the students in the United States actually finish college (The Organisation for Economic Co-operation and Development [OECD], 2010).

However, the demands of our economy are not changing. Currently, 44% of dropouts under the age of 24 are jobless. For most graduates to be prepared to earn a respectable wage in the United States, they will need some postsecondary education (Reich, 1996). Sixty-three percent of the jobs in the United States will require a college degree in the next decade (Carnevale, Smith, & Strohl, 2010). However, 75% of employers hiring new employees with 4-year college degrees reported that new hires

lacked basic knowledge and skills (The Conference Board, Corporate Voices for Working Families, The Partnership for 21st Century Skills, and the Society for Human Resource Management, 2006).

The recent statistics highlighting such negative aspects of the educational system in the United States have focused attention on reforming processes and policies.

Attention has turned to teachers to facilitate that change as research has consistently shown the quality of a teacher has been identified as the single most important school-based factor in student achievement (McCaffrey, Lockwood, Koretz, & Hamilton, 2003).

Statement of the Problem

Although research shows teachers have the greatest impact on student achievement (McCaffrey et al., 2003), determining the characteristics that define quality teachers and measuring these evidences that would capture effectiveness still remain problems in education (Partee, 2012). Researchers claim that although many theories and ideas about evaluation are recommended, there is no single instrument that quickly and accurately identifies and assesses teacher effectiveness (American Federation of Teachers, 2011). There is an expressed need for teachers and stakeholders to develop a shared understanding of good practice (Danielson, 1996).

Conceptual Framework for the Study

The mission of the Department of Public Instruction (DPI, 2012b) in a southeastern state is “every public school student will graduate from high school globally competitive for work and postsecondary education and prepared for life in the 21st century” (p. 1). As research has consistently shown, the quality of a teacher has been identified as the single most important school-based factor in student achievement (McCaffrey et al., 2003). In order to successfully accomplish the mission, the state has

built a foundation on that research and redeveloped a shared vision of school leadership and skills that teachers must use in order to obtain that goal and contribute to student success (DPI, 2012b).

As part of that vision, the state has implemented a new teacher evaluation system that directly applies to the district that was the focus of this study. Teachers are evaluated multiple times throughout the school year by administrators. The administrators utilize a state teacher evaluation instrument that measures teachers on five standards: leadership, learning environment, content knowledge, learning facilitation and teaching, and reflection. The instrument is completed by the school administration based on a series of lesson observations and classroom visits and by evaluating teacher involvement through the school and district (see Appendix A). Teachers are also given an opportunity to self-assess their performance (DPI, 2013d).

The evaluation instrument used by this district is based on the state professional teaching standards and frameworks developing 21st century learners. It is designed to support teachers in becoming effective leaders while encouraging quality teaching and student learning. The goal of the instrument is to foster an enhancement of professional practice in order to improve instruction, which has been directly linked to student achievement (McCaffrey et al., 2003). In addition, the instrument and implementation process is also designed to encourage professional growth as it encourages the establishment of professional goals and identification of professional development needs (DPI, 2012a).

As a component of the national No Child Left Behind (NCLB) initiative, states were required to implement standardized tests to increase teacher accountability for student achievement. Schools administer the standardized tests to measure the Annual

Yearly Progress (AYP) made by the students (NCLB, 2001). At the end of the school year, after standardized summative assessments are administered, teachers are given an effectiveness score based on a value-added model implemented by the state. This effectiveness score populates standard six on the state teacher evaluation tool. The value-added model varies from state to state and for the purpose of this study is represented by Education Value-Added Assessment System (EVAAS) data which are further described in Chapter 3. The value-added model is a statistical calculation that measures student growth based on the difference between students' predicted test scores and their actual performance on summative, standardized assessments. Students' previous test scores are used to determine predicted scores (DPI, 2013c).

On the state teacher evaluation instrument, the value-added score populates the sixth standard. Data are input each year to reflect the corresponding year evaluation. The purpose of the standard is to measure the work of the teacher based on the standardized test results. The state considers this measure to be an acceptable measure of progress for students based on established performance expectations using appropriate data to demonstrate growth (DPI, 2012d).

For fourth- through twelfth-grade teachers in tested subject areas during the 2012-2013 school year, the value-added score is a weighted measure where 70% is based on the student growth value for the individual students taught by the educator and 30% is based on the student growth value for the entire school. Based on the rating index score, a teacher is given a rating of Does Not Meet Expected Growth, Meets Expected Growth, or Exceeds Expected Growth. All local school boards are required to use student growth values generated through a method approved by the state board of education. An overall effectiveness rating of needs improvement, effective, or highly effective is assigned to

each teacher based on the rating by the value-added model and on the teacher evaluation instrument (DPI, 2013a). This process represents the evolution of how teachers are evaluated, as they are no longer evaluated solely on administrator observations. The new value-added initiative measures the effect a given teacher has on student growth. The two instruments measure different aspects of teacher performance.

Purpose of the Study

To date, there is a limited body of research that investigates the relationship between the two measures, as it is a relatively new topic. The purpose of this study was to determine the relationship between value-added data and the state teacher evaluation instrument among teachers of tested subjects in Grades 4-12 in a large urban school district in the aforementioned southern state.

As teachers have previously been evaluated based solely on administrator observations, the new value-added initiative will provide data on the effect a given teacher has on student growth. Since the two instruments measure different aspects of teacher performance, the researcher hypothesized that there would be a discrepancy between the score generated from a value-added model and the ratings given by administrators on the state-implemented teacher evaluation instrument.

Research Questions

1. What is the relationship between a teacher's effectiveness rating on the state teacher evaluation instrument and the rating determined from value-added data?
2. What is the relationship between a teacher's score for leadership as defined by the state teacher evaluation instrument and the rating determined from value-added data?

3. What is the relationship between a teacher's score for establishment of a respectful environment for a diverse population of students as defined by the state teacher evaluation instrument and the rating determined from value-added data?
4. What is the relationship between a teacher's score for knowledge of the content they teach as defined by the state teacher evaluation instrument and the rating determined from value-added data?
5. What is the relationship between a teacher's score for ability to facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data?
6. What is the relationship between a teacher's score for reflection on his/her practice as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Definition of Terms

Professional standards. A set of skills, knowledge, and behaviors that should be displayed by individuals in specific roles in state public schools (National Board for Professional Teaching Standards [NBPTS], 2013).

Teacher evaluation. Instruments and processes used to determine proficiency, measure educator performance against established standards, and help educators develop skills and knowledge based on observations conducted by administrators (DPI, 2012d).

Educator effectiveness system. State-developed evaluations of educators are completed through the use of an online tool which streamlines the process and facilitates the use of data at each level (DPI, 2013d).

Value-added model. Statistical calculation of the impact of a teacher or school

executive on the learning of students based on student performance, specifically growth, on standardized tests (DPI, 2013c).

Summary

In the United States, there is concern that students are not prepared to succeed in the 21st century economy. As the teacher has the greatest impact on student achievement, educational agencies and districts have focused on improving teacher performance (McCaffrey et al., 2003). The characteristics that define quality teachers are currently being identified, and evaluation systems are being developed to assess teacher performance around standards to aid in that process (Partee, 2012). Currently, the definition of teacher effectiveness for the district focused on in this study is determined by the state.

In a southeastern state in the United States, efforts to develop 21st century learners are driving the development of measures that evaluate teacher effectiveness. Currently, the state has chosen two methods to measure teacher performance. The teacher evaluation instrument rates teachers on five standards based on principal observations and school contributions. The value-added model identifies teacher effectiveness by measuring the impact a teacher has on students based on student achievement on standardized tests (DPI, 2013c).

Since the two measures evaluate teacher effectiveness through different methods, this study was conducted to determine the relationship between those two measures within the context of a single school district. The value-added data and the state teacher evaluation instrument data were analyzed among fourth- through twelfth-grade teachers of tested subjects in a large urban school district to determine the relationship between the ratings provided by each of the measures. Through this research study, the researcher

aimed to identify the current state of teacher evaluation and value-added ratings in the district, as well as to determine the extent of the relationship between these two measures.

Chapter 2: Literature Review

Introduction

As part of many evaluation systems, administrators visit classrooms to observe teachers delivering instruction. Typically, administrators note learning objectives, lesson strategies, learning environment, classroom leadership and management, student engagement, and student mastery of objectives. The administrator then uses that data to evaluate a teacher based on a set of skills, knowledge, or competencies. Administrators then typically meet with the observed teacher to discuss the lesson, evaluation ratings, and overall performance (Marzano, Frontier, & Livingston, 2011).

Research explains that teacher observations should be based on clear standards and should give administrators an accurate understanding of employee strengths and weaknesses. On most teacher observation evaluation forms, teachers are categorized by four to five ratings to identify differences in performance and effectiveness (The New Teacher Project [TNTP], 2009).

A value-added model is a statistical model used as a method of teacher evaluation that measures the teacher's contribution in a given year by comparing the current test scores of their students to the scores of those same students in previous years. The goal of using value-added models to evaluate teacher effectiveness is to allow educators an opportunity to compare teachers in terms of how much content their students learn each year, regardless of student characteristics. The scores indicate the amount of learning or improvement students made from 1 year to the next. There is no single value-added model currently used in teacher evaluations, but all of the models use prior student test scores to make estimated predictions about student performance and then measure the difference in the actual performance from the predicted score (Value-added teacher

evaluation, 2011).

Indicators of teacher effectiveness are needed for analysis to improve student achievement (Danielson, 1996). Many have been explored and discussed by researchers. Two of the most popular have proven to be classroom observation and student achievement data. However, principal-based observations and value-added measures both come with current policy conflicts and controversies, raising questions about how well evaluators are able to assess teacher effectiveness with observation and value-added frameworks. Researchers continue to examine how these measures determine aspects, components, and assumptions necessary to have a fair evaluation tool (Guarino, Reckase, & Wooldridge, 2011).

As theories around evaluating teacher effectiveness continue to change and evolve, there seems to be a growing necessity for systems to utilize a value-added model to measure that effectiveness. As researchers are finding that value-added scores are more objective in comparing teachers and measuring the effect teachers have on student growth, many systems continue to supplement their current teacher observation evaluations with value-added scores (Herman, Heritage, & Goldschmidt, 2011).

History of Teacher Evaluation

In the 1700s, local governments and clergy were responsible for hiring teachers and evaluating their teaching (Tracy, 1995). Individual supervisors or committees would establish criteria for teaching and monitor the quality of instruction (Burke & Krey, 2005). However, feedback varied because there was no standardization of their values (Marzano et al., 2011).

As industry began to rise in the 1800s, urban areas developed more complex school systems that required more specialized teachers and more knowledgeable

supervisors. Clergy were no longer considered suitable for those roles, since they lacked a broad understanding of content and skills (Tracy, 1995). Throughout the 1800s, stakeholders determined that teaching was an intricate profession and teachers required specific feedback. As a result, supervision began to shift focus to improving instruction (Blumberg, 1985). Although there was still no agreement on the necessary skills, education began to recognize that pedagogical awareness was a critical characteristic of effective teaching (Marzano, Waters, & McNulty, 2005).

By the 1900s, additional views of education began to emerge. As theory around learning and productivity developed, teacher evaluations began to reflect those values. The two dominant perspectives were those of John Dewey and Frederick Taylor. Dewey (1981) felt that schools should be organized around democracy, allowing students an opportunity to play an active role in their education by practicing citizenship. He pioneered progressive ideas such as student-centered learning, real-world connections, differentiation, and content integration (Dewey, 1981).

Taylor held a scientific view of management which began to compare school effectiveness to the productivity measures in factories. Taylor advocated for the measurement of specific behaviors, like those of factory workers, that would most improve production. He believed that there was one best method to completing any given task, and as business owners began adopting his principles, that belief was also translated to education (Marzano et al., 2011).

Leaders such as educational psychologist Edward Thorndike began to view Taylor's ideas of measurement as a tool for evaluating schooling and teachers. Cubberley (1929) applied those principles and described how schools can be managed as factories are, with children being the raw products that are developed to meet demands. His book

presented principles for school administrators to use to evaluate teachers and schools for productivity based on specific scales. The book also provided specific feedback for an administrator to provide a teacher based on ratings (Cubberley, 1929).

Wetzel (1929) built on Cubberley's (1929) work by proposing that administrators use measures of student learning to evaluate effectiveness. He suggested that student aptitude tests, measureable objectives, and measures of student learning be used by administrators to evaluate teacher effectiveness (Wetzel, 1929). As the scientific approach to education continued developing, a greater dependence on standardized tests also developed. Cubberley and Wetzel recommended that data be collected to provide feedback and further inform school decisions, whereas Dewey continued to focus on the goal of education (Marzano et al., 2011).

After World War II, evaluation began to shift the focus to the teacher as an individual, rather than using the scientific approach. Administrators were encouraged to help the teacher develop professional skills and manage personal needs. In January 1946, several articles were published in an issue of the *Educational Leadership* magazine that described a shift in teacher supervision. Coleman (1945) explained the importance of teachers being treated as people with personal circumstances and needs. Lewis and Leps (1946) wrote an article in the same publication outlining an evaluation process that included democratic ideals, initiative opportunities, human limitations, shared decision making, and delegation of responsibilities.

The role of supervisors during this time was still focused on managing the school, with a small aspect of that including teacher observation and evaluation. Melchoir (1950) described a variety of supervision responsibilities that included maintenance, curriculum, personnel, resources, materials, auxiliary services, and business and social events, in

addition to classroom observation. With the broad scope of the administrator responsibilities, there was still an understanding that teacher observation was an important part of improving instruction (Melchoir, 1950).

Several areas of supervision began being developed as a result of Melchoir's (1950) publication. Whitehead (1952) outlined six areas for administrators to focus on, bringing attention to the fact that supervision needed to work towards improving effective teaching. He expressed the need to improve the process for providing feedback for teachers and encouraged follow up via conferences (Whitehead, 1952).

The next major shift in educational evaluations was the development of clinical supervision models. The movement began in the 1950s and by 1980 about 90% of school administrators were using a clinical supervision model (Bruce & Hoehn, 1980). Cogan (1973) aided in the development of the "cycle of clinical supervision," a process for working with student teachers. The model evolved to a process of observation and discussion that allowed administrators and teachers to focus on growth and effectiveness (Goldhammer, Anderson, & Krajewski, 1993). There were five phases developed in the new model: preobservation conference, classroom observation, analysis, supervision conference, and analysis of the analysis. The work was further developed to include specific observable behaviors that guided administrator observations. The phases created a structure for which teacher evaluations were completed. They were initially developed to aid in the facilitation of critical conversations that allowed administrators to provide feedback but became viewed more as compliance tasks than as effective practice (Marzano et al., 2011). Through the development of clinical supervision models, characteristics of effective teaching still remained undefined (Marzano et al., 2011).

Throughout observation processes, teachers were being evaluated based on skills

and attitudes, by observation of teachers and students, or by combining several of those techniques. Many districts had a variety of summative and formative systems in place that were developed around teacher observations (Millard, 1976).

Peterson explored the importance of utilizing a variety of information to determine a teacher's true effectiveness. Data were collected regarding recognizing teacher credentials, personal characteristics, student outcomes, classroom visits, self-reports, student reports, peer review, competency-based teacher evaluation, and systematic observation. In many instances, he suggested that school systems utilize a combination of these techniques to successfully and appropriately evaluate teachers (Peterson & Kauchak, 1982).

As the work surrounding teacher evaluation progressed, the work of Hunter (1984) began to emerge. She created the seven-step model of a lesson that included anticipatory set, objective and purpose, input, modeling, checking for understanding, guided practice, and independent practice. Her model, known as mastery teaching, became used to evaluate the components included in teacher lessons. The model outlined the content that administrators expected to discuss and observe as they evaluated their teachers and provided feedback (Fehr, 2001).

As clinical supervision and mastery teaching continued to be used to supervise teachers, several different opinions began to emerge. One included focusing supervision on career goals and individual teacher needs (Glatthorn, 1984). Others focused attention on creating options to evaluate teachers differently based on experience (McGreal, 1983). One additional approach involved providing direct assistance to teachers and then giving them opportunities for participation in professional development and action research activities (Glickman, Gordon, & Ross-Gordon, 1998).

With the onset of these differing ideas, the RAND group conducted a study of 32 school districts to determine which evaluation practices were being used across the United States, finding the models being used were typically very standardized. Researchers found that models based on development and reflection were not specific enough for teachers to improve their instruction, and teachers preferred more formulaic processes. The study also uncovered four challenges with evaluation: lack of administration knowledge, teacher resistance to feedback, lack of uniform processes, and lack of training for evaluators. The RAND group responded to those challenges with 12 recommendations. The recommendations charged school districts with specific tasks to align goals of evaluation to organizational goals; provide time, resources, and training to evaluate accurately and effectively; and to monitor the quality of evaluators, including expert teachers in the process when possible (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984).

Danielson (1996) published her framework for teaching which outlined four categories of instruction. She included competencies, knowledge, and skills of educators in her Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities domains. She identified 76 characteristics of quality teaching and included four levels of performance ranging from Unsatisfactory to Distinguished. Danielson felt the model, built on research, was able to be used across a variety of levels and disciplines to capture the complexity of teaching, create a language for conversation, and provide a structure for reflection. The model outlined the most specific and detailed approach to evaluation in education (Danielson, 1996).

In the 21st century, education has seen yet another shift in supervision and evaluation. Practice has moved from observing teacher behavior to noting student

achievement as a measure of teacher performance (Marzano et al., 2011). Tucker and Stronge (2005) began to draw even more attention to linking teacher evaluation and student learning. They recommended that data surrounding student gains in learning should be used to determine a teacher's effectiveness in addition to traditional observations. They studied four school districts across the United States that used both measures to evaluate effectiveness. They supported the use of both measures, but strongly suggested that student achievement be used to provide feedback about effectiveness (Tucker & Stronge, 2005).

As focus began shifting to student performance, value-added measures of teacher effectiveness began to emerge to try to identify effective schools, classrooms, and teachers. In their study, Weerasinghe, Orsak, and Mendro (1997) recommended that data be collected and reported using statistically valid models and by comparing pretest and posttest scores and student gain scores. As Tucker and Stronge (2005) advocated, research continued to suggest that a more accurate evaluation could be compiled using a variety of data (Weerasinghe et al., 1997).

There were many documented practices utilized in districts across the United States for evaluating teacher performance. Still, the common practice included principal observation to provide an overall representation of classroom practices, content capacity, and pedagogical knowledge. However, stakeholders began to view this measure as a subjective assessment of teacher competence. School districts began exploring additional measures used to evaluate teacher performance. Researchers suggested that the utilization of value-added models for teacher evaluation was one solution for the complaint that observations were too subjective. Districts started looking at how to measure student achievement gains (Alicias, 2005). Most commonly, student pre and

posttest data began being compiled, so growth could be measured using a value-added analysis model. Specifically, in Tennessee, the Value-Added Assessment System was developed by Dr. William Sanders. His findings reported that the models were able to identify assumptions and factors of student achievement as a result of implementation (Sanders & Horn, 1994).

A research study of the validity of teacher evaluation was conducted in four cities across the United States: Cincinnati, Ohio; Los Angeles, California; Reno, Nevada; and Coventry, Rhode Island. The teacher evaluation instruments included in the study were all developed based on the Framework for Teaching created by Danielson (1996). In addition to the instrument, researchers used a value-added model to assess teacher contribution to student achievement. Researchers correlated the evaluation scores with the value-added findings. The student-reported findings ranging between $r_s=0.37$ and $r_s=0.11$, with the strongest correlations found in California and Nevada. The stronger correlations were attributed to the use of multiple evaluators; strong instructional culture; and intensive, high-quality training (Heneman, Milanowski, Kimball, & Odden, 2006).

Researchers hypothesized that they would find a relationship between what teachers were observed doing and the achievement data, as calculated by utilizing a value-added model. There was some positive correlation found; however, the findings were not consistent for each location. Researchers determined as a result of the study that value-added data cannot be used as a substitute for classroom observation data to measure teacher effectiveness.

Recent Changes to Teacher Evaluation

In 2009, *The Widget Effect*, a study of teacher evaluations, criticized current practices in the United States claiming systems were not providing accurate feedback and

evaluation of teachers (TNTP, 2009). The research evaluated practices in 12 districts in four states and included approximately 15,000 teachers; 1,300 administrators; and 80 education officials. School districts in Arkansas, Colorado, Illinois, and Ohio varied in size, geographic location, evaluation practices, and approach to teacher performance management. The smallest district included in the study was Jonesboro Public Schools, serving approximately 4,450 students; and the largest, Chicago Public schools, serves nearly 413,700. Teachers and administrators were engaged through surveys; and officials, leaders, policymakers, and advocates participated in advisory teams and provided input about the study design, data collection, and recommendations (TNTP, 2009).

TNTP identified *The Widget Effect* as the description of the current trend among school districts that assumes teachers are similar in their classroom effectiveness, whereas individual strengths and weaknesses are not recognized or identified. TNTP (2009) reported that teacher evaluations are not currently distinguishing “great teaching from good, good from fair, and fair from poor” (p. 6). As researchers reviewed teacher observation data in the 12 districts, they found that between 94% and 99% of teachers are receiving the highest ratings; and less than 1% of teachers receive unsatisfactory ratings. However, 81% of administrators and 57% of teachers recognized that they were aware of tenured teachers who were performing poorly in their respective schools (TNTP, 2009).

TNTP (2009) found that the short and infrequent classroom observations have not accurately measured or recorded teacher effectiveness, as they are largely influenced by relationships and teacher expectations. Seventy-three percent of teachers responded that they were not provided with areas to improve, and only 45% reported receiving feedback that was used to improve their practice. Similarly, 59% of teachers and 63% of

administrators felt that school districts were not effectively identifying, compensating, promoting, and retaining the most effective teachers. Overall, *The Widget Effect* summary claimed that school districts are indifferent to instructional effectiveness based on their findings (TNTP, 2009).

As a result of the findings, the report recommended a reform of current evaluation practices. TNTP (2009) identified four actions for school districts to take to ultimately reverse *The Widget Effect* in schools. Initially, school districts were charged with implementing comprehensive systems that “fairly, accurately and credibly differentiate teachers based on their effectiveness in promoting student achievement” (TNTP, 2009, p. 27). Additionally, reports recommended training and accountability for administrators, integrated human capital policies, and more efficient dismissal policies (TNTP, 2009).

Many states, including Pennsylvania, began working to develop a new statewide teacher and administrator evaluation system based on *The Widget Effect* report. The Team Pennsylvania Foundation piloted a new evaluation with the aid of a stakeholder group that included leaders from the state department of education, state education association, school districts, and the community. The pilot developed and implemented performance standards in order to improve classroom observations and student achievement. In addition, value-added models were developed to estimate teacher effectiveness (Lipscomb, Chiang, & Gill, 2012). Other states like Ohio also began working to improve teacher effectiveness using multiple strategies and measures such as professional development experiences for teachers (Lloyd, 2009).

Multiple research studies have since been conducted deeming teacher observations by administrators a success when processes were reformed. For example, a 2-year Excellence in Teaching Pilot was conducted in Chicago that utilized observation

techniques to assess teacher performance. The pilot was developed to increase student performance by providing teachers with feedback about their instructional practices. The pilot involved identifying teacher strengths and weaknesses through the use of principal observations of teaching practice. The observations were conducted twice a year using the Charlotte Danielson Framework for Teaching. Postobservation conferences between the principal and the teacher were also outlined to discuss evaluation results. The study had broad implications for districts and states nationwide as they continued trying to redevelop evaluation systems that rely on classroom observations. Overall, the research identified that the process was a success when the policies and guidelines were followed strictly (Sartain, Stoelinga, & Brown, 2011).

However, as pressure has been placed on districts to reform their evaluation systems, value-added models continue to emerge as a way to measure student gains. Value-added models assign statistically valid expectations for student achievement based on prior performance and demographics, which allows teacher performance to be measured against the expectation which removes external factors (Nicholson & Brown, 2010). Value-added models of measuring student achievement have become increasingly more popular among school systems to evaluate, promote, compensate, and dismiss teachers based on student test scores. The method is considered objective in comparison to many of the other techniques of evaluating performance. The model provides a statistical measure that is being used to assess teacher quality (Corcoran, 2010).

The recent interest in tying student performance to teacher evaluation by utilizing a value-added method is increasing across districts. However, scores and judgments of teacher effectiveness still vary across statistical models, classes taught, years, and student characteristics. Value-added analyses are used not only for teacher evaluations but also

as a means to inform decisions about school roles, professional learning community effectiveness, student aptitudes, and home efforts (Newton, Darling-Hammond, Haertel, & Thomas, 2010). Although multiple measures are beginning to be considered, there is still no uniform method for evaluating teachers.

As education continues to evolve, researchers have compared traditional approaches to measuring achievement in high-stakes testing environments. Previously, data have been limited to the static snapshots of student performance, highlighting student status on one test, given one period of time. However, value-added models provide more insight about the effect a teacher has on students than those traditional proficiency analyses as value-added models link grades and courses to create a longitudinal measure of student performance (O'Malley, McClarty, Magda, & Burling, 2011). Utilizing value-added models to compare data, calculations show student growth from year to year instead of patterns in performance as it pertains to proficiency. Assessment data provided by value-added models have the potential to positively affect academic performance and motivation (Anderman, Anderman, Yough, & Gimbert, 2010). Value-added models have been found to collect more direct evidence from student data, enhancing the ability to draw inferences about student growth. Those concepts have been expanded to inform evaluations of teachers (O'Malley et al., 2011). Recently, the United States Department of Education supported those claims, stating “instead of a single snapshot, we will recognize progress and growth” (Duncan & Martin, 2010, p. 2).

Researchers also suggest that data from value-added models can be compared with teacher observation assessments also in use across the country (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). As a result, many districts are

implementing the use of value-added models with their teacher traditional evaluation systems. For example, in the District of Columbia Public Schools, a value-added model has been developed by school system officials, Eric Hanushek of the Hoover Institution at Stanford University and Tim Sass of Florida State University, to measure teacher effectiveness. Teachers are currently evaluated by their value-added scores and teacher observation results on student achievement, instructional expertise, collaboration, and professionalism (Isenberg & Hock, 2011). As many states and districts are currently developing student growth metrics for teacher evaluation purposes, there is still relatively limited evidence to guide development. Districts in states such as Delaware, Tennessee and Ohio, as well Dallas Independent School Districts and Tulsa Public Schools, are working to better understand the value-added measures in order to identify ways to match data and context for their teachers. Educators, administrators, and policymakers have expressed that these measures are necessary to improve student learning, provide professional development, and increase accountability for teachers as they continue preparing students (O'Malley et al., 2011).

Similarly, the Los Angeles Unified School District has also implemented the use of a value-added model to track student growth over the school year. The analyses have evaluated the effects of teachers on student performance as evidenced by California standardized test scores. Teachers were able to be classified into levels of effectiveness based on the ratings. In 2010, the Los Angeles Times even published the value-added results from student test data to provide information about schools and teachers in the school district (Briggs & Domingue, 2011).

The Measures of Effective Teaching (MET) project funded by the Bill and Melinda Gates Foundation (2013) found that “it is possible to develop reliable measures

that identify great teaching” (para. 3). During the 3-year study, researchers aimed to identify and promote effective teaching practices by collaborating with almost 3,000 teacher volunteers from seven public school districts across the United States. Teachers volunteered from Charlotte-Mecklenburg Schools, Dallas Independent Schools, Denver Public Schools, Hillsborough County Public Schools, Memphis Public Schools, New York City Schools, and Pittsburgh Public Schools. Researchers from dozens of independent teams gathered to investigate ways to identify and develop effective teaching. Research partners included American Institutes for Research, Cambridge Education, University of Chicago, The Danielson Group, Dartmouth College, Educational Testing Service, Empirical Education, Harvard University, NBPTS, National Math and Science Initiative, New Teacher Center, University of Michigan, RAND, Rutgers University, University of Southern California, Stanford University, Teachscape, University of Texas, University of Virginia, University of Washington, and Westat. Through the development of their process of Measuring Effective Teaching, Ensuring High-Quality Data, and Investing in Improvement, they found that effective teaching could be identified using a combination of three measures: classroom observations, student surveys, and student achievement gains. The MET project utilized those three key assessments to measure teachers. They utilized data from student surveys to measure the supportiveness of the classroom, teacher observation to assess practice, and student assessment data to measure student growth (Bill and Melinda Gates Foundation, 2013).

Researchers also stressed that weights of the measures should be balanced to prevent focus on any single measure and neglect of another. The report generated from the study suggested that school districts adhere to more specific guidelines when conducting classroom observations to ensure validity, reliability, and accuracy. As a part

of those guidelines, the report recommended that evaluators undergo rigorous training and assessment on differentiating performance, observations should be conducted by more than one observer, and a variety of lessons be observed (Bill and Melinda Gates Foundation, 2013).

The first recommendation to school districts was to balance the weights of the multiple evaluation aspects identified around effective teaching. In the MET study, researchers recognized that the measures that were weighted the highest were valued the most. It was suggested that between 33% and 50% of a teacher's overall rating be determined by student growth and achievement measures. Maintaining the balance ensures that there is not a narrow focus on any single measure but encourages improvement on all measures (Bill and Melinda Gates Foundation, 2013).

Additionally, the MET study researchers recommended that evaluators participate in rigorous training to improve validity and accuracy but also to prioritize support and feedback. In the study, Hillsborough County had two types of observers participate: administrators and peer observers. The researchers specifically noted that all trainers had gone through professional development to learn about the process and how to appropriately utilize the tool, finding observations were more consistent and ratings were similar when compared with a variety of observers (Bill and Melinda Gates Foundation, 2013).

In reviewing approximately 300 evaluations conducted by the observers, the MET study found that reliability increased when observations were completed by more than one person and a variety of lessons were observed. The highest reliability was found when adding multiple observers and multiple lessons of a single teacher. In this specific study, reliability increased from .51 to .58 (based on a 0 to 1 scale) when a second lesson

was observed by the same observer who completed the initial observation. However, reliability increased more than twice as much to .67 when the two observed lessons were completed by two different observers (Bill and Melinda Gates Foundation, 2013).

The researchers concluded the findings with the fact that the MET process makes meaningful distinctions between teachers. Where traditional evaluation systems have suffered from *The Widget Effect*, the MET project data suggested that teacher effectiveness did not group high percentages as satisfactory nor did it distribute performance equally. Contrastingly, 50% of the teachers scored within 0.4 points of each other on a four-point scale and bunched at the center of distribution. Less than 10% of teachers scored in the lowest and highest percentages. The data suggest a large middle category of effectiveness. As a result, researchers suggested that districts continue developing their evaluation systems to align with the Common Core State standards to help all teachers improve their practice and the outcomes for students (Bill and Melinda Gates Foundation, 2013).

Although many systems are implementing value-added models to measure student growth and achievement, there have been several complaints around using the models. Additional research has been conducted which focuses on the report regarding value-added models for teacher evaluation. Reporting seems to suffer from the “Lake Wobegon effect,” wherein, similar to the “Widget Effect,” all teachers are still being rated above average. As the evaluations focus on consequences for teachers instead of consequences for students, administrators are still accused to trying to protect their teachers from misclassification (Value-added teacher evaluation, 2011).

State Changes in Evaluation

Race to the Top. The Race to the Top initiative was a \$4.35 billion grant created

by the United States Department of Education to encourage states to reform education. Funded by the ED Recovery Act, President Obama and Secretary of Education Arne Duncan announced the contest in July 2009. States were encouraged to apply for the grant if they satisfied specific educational policies. They have been able to utilize the funds to build performance-based standards for teachers and principals, comply with Common Core standards, lift caps on charter schools, turn around low-performing schools, and build data systems (American Recovery and Reinvestment Act of 2009 [ARRA], 2009).

The southeastern state where this study is held was one state to receive grant money in 2010, receiving nearly \$400 million. The state public school system has worked to remodel educational approaches through a new initiative with new standards and a new accountability model for educators to prepare students for college and careers. The initiative strives to increase student achievement, close achievement gaps, and increase the number of career- and college-ready graduates by ensuring quality teachers are in every classroom. The state plan sets goals to create strong leaders, a fair evaluation system, tools and training to improve practice, an improved supply of teachers, and support for low-performing schools.

As a result, the state developed a new evaluation system to ensure every student was given the opportunity to grow academically, be held to high academic standards, and graduate college- or career-ready by having an effective teacher at a school with an effective principal (DPI, 2013e).

Educator evaluation system. The educator evaluation system developed in the state is based on the Framework for 21st Century Learning and includes professional standards and evaluation processes for every educator in the state. The Framework for

21st Century Learning was developed by the Partnership for 21st Century Skills (P21, 2011), a coalition of business, education, and policy leaders founded in 2002. The United States Department of Education, AOL Time Warner Foundation, Apple Computer, Cable in the Classroom, Cisco Systems, Dell Computer Corporation, Microsoft Corporation, and the National Education Association were instrumental in the development of P21 (P21, 2011).

Data for the educator evaluation system is captured annually in an online tool, mostly through observations conducted by administrators. The purpose of the educator evaluation system is to promote leadership, quality teaching, and student learning through a fair growth model. The evaluation process is based on gathering information from multiple data sources, employee artifacts, and other evidences to measure employee performance and effectiveness. One component of the system is that employees are responsible for setting performance goals as they strive to grow through professional development opportunities.

The state Professional Teaching Standards Commission has defined the roles and responsibilities that teachers need to fulfill in order to successfully educate students in the 21st century. Those definitions have been used to create the foundation for the educator evaluation system, the rubric for evaluating teachers across the state. They have outlined roles around leadership, classroom environment, content knowledge and delivery, learning facilitation, and reflection. Those areas were used to develop five evaluative standards for teachers (DPI, 2012d).

Professional standards were developed for all employees of the state's public school system. Professional organizations, staff members and representatives from higher education were able to provide input as the standards were written. The standards

include the skills, knowledge, and behaviors expected for each given role (DPI, 2012e).

The standards contain information for evaluators about conducting role-specific evaluations and providing appropriate feedback. Processes were also developed to allow evaluators to facilitate learning and growth among their employees, as well as measure educator performance against the standards. The evaluation contains five different ratings: Not Demonstrated, Developing, Proficient, Accomplished, and Distinguished. The state expectation is that every professional should obtain a rating of at least proficient on the standards (DPI, 2013d).

Standard one: Teachers demonstrate leadership. According to the state Professional Teaching Standards Commission, one critical aspect of educators is that they participate in actively leading the school in collaboration with the administration. Teachers have set expectations to lead in their classrooms, school, and in the teaching profession (DPI, 2013d).

In their classrooms, teachers are responsible for the learning and progress of all students. They must contribute to and impart the vision to students that they graduate from high school ready to pursue postsecondary education and compete in a global workforce. Teachers are responsible for communicating those ideals to students to ensure they are prepared for the 21st century. To do that, teachers are expected to utilize a variety of data sources to aid in the setting and development of goals that meet the needs of each student. As the year progresses, teachers are responsible for evaluating student progress towards those goals and making adjustments as necessary. As part of leading their classrooms, teachers must establish a classroom culture that empowers students and develops lifelong learners. Teachers who lead in their classrooms are responsible for student learning, communicating the educational vision to students,

planning and setting goals, utilizing assessment data to monitor progress, and empowering students (DPI, 2012e).

Teachers are also responsible for leading within the school. Teachers are expected to work collaboratively with administrators and peers to create professional learning communities in the school. Professional learning communities facilitate conversations around data analysis and the development of improvement strategies to enhance student learning and teacher working conditions. Professional learning communities work to develop and support the school improvement plan as well as the school budget and professional development opportunities for teachers. As school leaders, teachers are also afforded the opportunity to participate in the hiring process and the supporting of new teachers and staff members through mentoring and coaching. In the school, teachers work together to improve the overall effectiveness of the school by participating in professional learning communities, analyzing data, developing goals and strategies in the school improvement plan, aiding in the development of the school budget and professional development, participating with hiring, and mentoring and supporting teachers (DPI, 2012e).

Lastly, teachers are evaluated on their impact on the teaching profession. Teachers have the responsibility of advocating for their students and their school by participating in educational decision-making processes. Teachers are responsible for their part in improving the profession by promoting professional growth for all educators. Additionally, teachers should advocate for positive changes in policies and practices. Teachers are evaluated on their ability to improve the education of students; for example, their participation in the implementation of new initiatives. As part of their leadership roles, teachers must demonstrate high ethical standards including honesty, integrity, and

respect (DPI, 2012e).

Standard two: Teachers establish a respectful environment for a diverse population of students. The second standard in the teacher evaluation instrument assesses the classroom environment created by the teacher. The foundation developed by the commission defines that a classroom should be positive and nurturing and one that facilitates the development of a caring relationship between teacher and student. Teachers are responsible for ensuring the learning environment is inviting, respectful, supportive, inclusive, and flexible. As a part of that responsibility, teachers also have to embrace the diversity around them. To demonstrate this, teachers should select materials and organize lessons to include contributions from a variety of cultures. In doing so, teachers must also recognize the influence of culture on a student's development, personality, and performance. Teachers must understand the effects of race, ethnicity, gender, religion, and other cultural influences and consider those perspectives and dispositions when planning instruction. Teachers who develop a respectful, diverse environment are knowledgeable of diverse cultures, counteract stereotypes and incorporate contributions, recognize cultural influences on a child, and consider different points of view (DPI, 2012e).

Another element of standard two refers to the treatment of students. The evaluation rubric outlines the high expectations that teachers maintain for all students, recognizing the appreciation of differences and contributions from each of them. Not only are teachers responsible for building those relationships with students but also treating their families as significant contributors to success. It is the role of the teacher to help improve communication and collaboration between the school and the community. Relationships are built to promote trust and understanding among all stakeholders. To

accomplish this, teachers must often seek solutions to aid families in overcoming obstacles that prohibit family involvement. Effective teachers treat students and families with respect to build positive relationships (DPI, 2012e).

Lastly, teachers are responsible for adapting their practices to promote learning for all students, including those with special needs. Teachers work collaboratively as leaders in their buildings and professional learning communities with specialists to meet those needs. Teachers collaborate with specialists and utilize best practices, specifically inclusion models, to engage students to meet student needs (DPI, 2012c).

Standard three: Teachers know the content they teach. In the third standard in the evaluation, teachers are assessed on their content knowledge and delivery. Teachers are expected to create engaging, relevant lessons that connect content to students' lives. Effective teachers integrate multiple disciplines and subjects throughout their lessons and activities. In addition to the subject content, teachers are also responsible for teaching 21st century content such as global awareness; civic literacy; financial literacy; health awareness; and a variety of higher-level skills such as critical thinking, problem solving, and technology literacy (DPI, 2012e).

All teachers in the state have a specific course of study to guide instruction. Teachers are assessed on the alignment of lessons and activities with those learning standards. Teachers collaborate with other staff to further understand and investigate those content standards as well as develop additional strategies to make the curriculum rigorous and relevant for all students (DPI, 2012e).

One of the critical components of content knowledge and delivery is that teachers make instruction relevant to students. Teachers have a responsibility to focus instruction on 21st century life skills such as leadership, ethics, accountability, adaptability, personal

productivity, personal responsibility, people skills, self-direction, and social responsibility in order to communicate the vision to ensure students graduate globally competitive. Teachers are fully responsible for facilitating the learning of core content, 21st century content, and the development of higher-level thinking and 21st century skills (DPI, 2012c).

Standard four: Teachers facilitate learning for their students. The expectation of the Professional Teaching Standards Commission is that the teacher is the facilitator of learning in the classroom. However, within that responsibility, teachers are expected to allow students opportunities to create their own learning experiences. Effective teachers accomplish this through innovative lessons and providing opportunities for students to collaborate and communicate with their peers (DPI, 2012e).

As instructional facilitators, teachers must understand the levels of development of their students including intellectual, physical, social, and emotional development, and the way students think and learn. As a result, teachers are able to differentiate learning for their students based on their specific strengths and needs. Teachers plan instructional lessons, activities, and opportunities using a variety of instructional methods and learning styles for their students based on the content standards and those student needs. Effective teachers made adjustments to instruction as needed based on those student needs and their understanding of research-based best practices (DPI, 2012e).

Additionally, teachers are expected to integrate and utilize technology to maximize student learning. Teachers use technological resources to promote critical thinking, problem solving, communication, innovation, collaboration while teaching appropriate usage, and the ability to identify credible sources (DPI, 2012e).

Teachers are responsible for encouraging students to ask questions, draw

conclusions, make complex choices, and synthesize information. To be globally competitive, it is important for students to understand how to work cooperatively in teams with their peers. Teachers are expected to provide opportunities for them to assume leadership roles, improve collaborative skills, and interact with people from different cultures and with different strengths (DPI, 2012e).

Another responsibility of teachers within the fourth standard is effective communication. It is critical that teachers are clearly understood by their students and that their ideas are clearly articulated. Effective teachers successfully disseminate information to their students (DPI, 2012c).

Standard five: Teachers reflect on their practice. The fifth standard outlines the expectation that teachers reflect on their practice. As aforementioned, teachers have a responsibility to continuously gather data about student learning through a variety of assessments; however, it is critical that teachers also continuously analyze that data and make appropriate changes to instruction as necessary. Teachers are expected to critically and systematically review data to understand learning in their classrooms to ultimately improve student performance (DPI, 2012e).

As part of their reflection, teachers are responsible for communicating the value of learning and growth with their students and developing their own professional growth. Effective teachers seek out professional development opportunities that support the success of their professional goals. Teachers are expected to participate in professional development that aligns with core and 21st century content, student and personal goals, and the needs of the students in their classrooms (DPI, 2012c).

Performance rating scale. Every teacher in the state is rated on their mastery of the aforementioned standards. The multiple evaluations conducted on the elements

previously described are used to determine the overall rating for each standard on the final summary evaluation at the end of the school year. On each of the five standards, teachers receive one of five ratings: Not Demonstrated, Developing, Proficient, Accomplished, or Distinguished. A rating of Not Demonstrated is earned if a teacher did not show growth towards a specific competency. If Not Demonstrated is used, the evaluator must provide a written comment as to how they looked for the element, but it was not observed throughout the course of the school year. A rating of Developing is earned when a teacher has demonstrated some growth towards achieving the standard but has not yet met expectations for those responsibilities. Proficient ratings are given when teachers demonstrate basic performance expectations outlined in the standard. Accomplished and Distinguished ratings are utilized for teachers who exceed basic competency on standards based on the frequency of that performance. Accomplished ratings are given when teachers exceed the expectations on the standard most of the time. Distinguished ratings are reserved for teachers who significantly exceed the basic expectations (DPI, 2012d).

Teacher and principal responsibilities. The last component outlined by the state Professional Teaching Standards Commission is responsibilities of the evaluation process for principals and teachers. Principals are responsible for understanding the teaching standards before they are able to accurately evaluate their direct reports. Each evaluator is required to participate in training about the standards and evaluation process; as they are ultimately responsible for supervising all evaluations to ensure a fair process is completed, all steps are followed appropriately, and all reports are accurate. As part of the evaluation process, administrators are accountable for developing their employees by identifying teacher strengths and weaknesses, making recommendations for

improvements, and developing and implementing action plans (DPI, 2013e).

Teachers also have responsibilities in the evaluation process. First, they must know the standards and the evaluation process. New teachers typically attend multiple training sessions to gain an understanding of the evaluation standards process. Similarly, experienced teachers are re-oriented to the process at the beginning of each school year. In the state, the orientation must be done within the first 10 days of the school year to ensure teachers are made aware before any observations or evaluations begin and to allow them appropriate preparation time for each component included. Throughout the school year, teachers are also responsible for gathering data and evidence to support their performance on the five standards. Those artifacts collected are shared with administrators during evaluation conversations and postobservation discussions. Teachers are expected to set goals and continuously develop strategies to aid in improving their performance and meeting those goals (DPI, 2013e).

As a result of *The Widget Effect* and Race to the Top requirements, the educator evaluation system has also been developed with a student growth component, standard six. The state links the impact of the teacher on student achievement to the evaluator observation (DPI, 2013c).

EVAAS. The state currently uses EVAAS, developed by a private statistics vendor, SAS, to compute teacher effectiveness. The rating generated by SAS populates standard six on the teacher evaluation tool. EVAAS became part of the educator evaluation system during the 2011-2012 school year after the state board and department of instruction reviewed various growth and value-added models. The state was seeking a metric that compared teacher impact across grades, subjects, schools, and districts. EVAAS provides information about past performance and student-predicted scores to

ultimately determine educational growth over the course of the school year (DPI, 2013a).

Reliability of the model was reviewed and confirmed by the Consortium for Educational Research and Evaluation, as well as WestEd (2012) research and policy development agency, four United States Department of Education peer review committees, the government accountability office, and the RAND Corporation (DPI, 2013b).

As a result of EVAAS being implemented as the statewide student achievement growth model, standard six was added to the evaluation for teachers and school administrators to measure growth when common assessments are administered across the state. Based on the index score, a teacher is provided with a rating of Does Not Meet Expected Growth, Meets Expected Growth, or Exceeds Expected Growth. EVAAS assesses the impact of teachers, schools, and districts on student learning. SAS provides districts with customized reports that predict student scores, show teacher impact, and assess subgroup performance (DPI, 2013a).

Recent District Initiatives

The federal Teacher Incentive Fund (TIF) began supporting the development of compensation systems in high-need schools based on teacher and principal performance. Robert Meyer and Michael Christian examined various TIF performance-pay plans. At a presentation at the National Center on Performance Incentives research to policy conference in February 2008, their analysis focused on the use of value-added models to evaluate teacher performance in TIF plans (Vanderbilt University, 2008).

In 2010, a large rural school district in a southeastern state began a Pay for Performance initiative. The school system used TIFs to support their work. The initiative began by identifying multiple measures of teacher effectiveness and defining

how those were applied in the classroom. The ultimate goal was to have all school district employees on a compensation plan that was based on multiple measures of effectiveness. The district wanted to directly align the compensation budget with student achievement. Because teachers have historically been paid due to years of experience, the system concluded that the top performing teachers were not necessarily being paid their worth (School District, May 5, 2010; School District, February 8, 2010).

During the 2010-2011 school year, teams of 10-12 stakeholders met to discuss the measures of effectiveness. The project prided itself on the fact that an invitation was sent to every teacher employed by the school district. The initial seven teams were Contributions to Schools, Value-Added, Observations, Student Surveys, Hard to Staff Schools and Subjects, Student Work, and one group was dedicated to exploring new options. All of the teams had multiple meetings, conducted research, held focus groups, created an update in May 2011, and presented their findings to the Board of Education (School District, 2012).

In September 2011, the initiative transitioned to the Teacher Effectiveness Project. The goal of the work shifted from developing a compensation plan to developing talent across the district to impact student achievement. The district aimed their design teams toward creating a more effective evaluation system and withdrew their objective of tying performance to compensation in an effort to create opportunities for highly effective teachers to increase their compensation. Teams, including district services, executive staff, school administrators, and instructional staff, met to continue the work on recommendations about teacher effectiveness from the previous school year. The 2011-2012 teams included Professional Learning Communities (previously known as Contribution to School), Value-Added, Teacher Observations, Student Surveys, Hard to

Staff Schools and Subjects, Teacher Work Product, Student Learning Objectives, Content Pedagogy (developed from the New Options team), and one group continued to facilitate discussion that explored other options. The other options group developed into Classroom Management through the course of the work. In May 2011, the groups presented their findings to the Board of Education and Executive Staff members (School District, March 6, 2012).

With the hire of a new superintendent and new state policies, the work of the Teacher Effectiveness project was transitioned once again. As a result, a teacher Compensation Task Force was built to lead the direction of the system. All of the facilitators of the design teams met with various school district and community stakeholders to focus the work back to developing a compensation plan that successfully identifies and recognizes effective teaching. Recommendations from the previous teacher working teams were reviewed and additional research was conducted. The team discussed several plans that had already been implemented in various states across the United States and used their findings to guide the work. The team aimed to create a more reliable program to evaluate teacher performance and to promote growth among teacher and student performance. The team presented their recommendations for an alternative compensation plan in July 2013 (School District, January 18, 2013).

The school district has also begun a new partnership with several community partners in an attempt to improve teacher compensation and overall performance. One project has resulted in the forming of a nonprofit organization partnering public and private institutions. The project operates as one zone within the school system, working directly with nine schools in the district. The goal of the program is to improve student achievement in high-need schools that have a history of poor performance by providing

additional supports for students and teachers while providing more competitive salaries to employees in those schools. The program involves extensive reading and math education while incorporating innovative learning strategies. The project focuses on time, talent, technology, and community engagement. To access the curriculum, learning opportunities are extended and students are provided access to technology. Schools work diligently to motivate parents and the community to engage with the school. Specific consideration is given to the traditional recruiting, training, and retention processes to place excellent leaders and effective teachers in each school. As a result, teachers are evaluated utilizing a variety of measures that directly impact teacher compensation. The project utilizes alternative compensation strategies based on student achievement growth (School District Project, 2013).

Summary

Throughout the course of history, teacher improvement and evaluation systems have focused on classroom observations. Traditionally, administrators visit classrooms to observe learning objectives, lesson strategies, learning environment, classroom leadership and management, student engagement, and student mastery of objectives. Administrators have used those observations to evaluate teachers based on identified skills, knowledge, or competencies (TNTP, 2009).

Most recently, there has been a shift in teacher evaluation to include student performance data to evaluate teacher impact and effectiveness. Value-added models are being used more frequently in evaluation systems across the country (National Council on Teacher Quality, 2013). Researchers continue to examine how these measures determine aspects, components, and assumptions necessary to have a fair evaluation tool (Guarino et al., 2011).

As evaluations striving to quantify teacher effectiveness continue to change and evolve, stakeholders understand the need for systems to include a value-added measurement in addition to traditional observation methods. Researchers continue to find objectivity in value-added scores, making them suitable for teacher comparison for and understanding the effect teachers have on student growth. Evaluation systems based on teacher observation are increasingly being supplemented by value-added scores (Herman et al., 2011).

Chapter 3: Methodology

Introduction

This correlational study strives to determine the relationship between two measures of teacher effectiveness in a southeastern state in the United States. Currently, the state uses a teacher evaluation instrument that rates teachers based on principal observations on five standards. The state has also recently implemented the use of a value-added model that measures a teacher's impact on student learning based on student achievement on standardized tests (DPI, 2013c).

This study was conducted in a single, large urban school district in the state to determine the relationship between these two measures within the context of a single school district. The value-added data and the state teacher evaluation instrument data were analyzed to determine if there was a correlation between the ratings provided by each of the measures.

As the teacher has the greatest impact on student achievement, educational agencies and districts have focused on improving teacher performance (McCaffrey et al., 2003). This study aimed to compare methods used to assess quality teaching.

Problem

The quality of a teacher is the most important school-based factor that impacts student achievement (McCaffrey et al., 2003). The challenge continues to be how to determine the characteristics that define quality teachers and how to identify measures that can be used to accurately evaluate educator effectiveness (Partee, 2012). There is currently not a single instrument that stakeholders agree would quickly and accurately assesses teacher effectiveness (American Federation of Teachers, 2011).

Purpose

The purpose of this study was to determine the relationship between a teacher's value-added score determined by student growth on standardized assessments and ratings given based on the state teacher evaluation instrument based on classroom observations among teachers in a large urban school district.

Population

District. The data collection took place in a large, urban district in a southeastern state. The school district is an historical staple in the city, with the first public schools opening in 1882. The vision of the district is “to provide all students the best education available anywhere, preparing every child to lead a rich and productive life;” with the mission being, “to maximize academic achievement by every student in every school.” To meet those needs, the school district currently operates on a \$1.2 billion budget, spending \$8,473 per pupil (School District, 2013b). The district ranks slightly above the state average per pupil expenditure of \$8,436 (DPI, 2013f). The district also has exceptional support from local corporations, faith communities, and businesses that provide 91,267 mentors and volunteers to support learning (School District, 2013b).

In 2011, the district won the Broad Prize for Urban Education for academic gains and narrowing achievement gaps. In high school math, students performing at the highest achievement level increased an average of six percentage points per year between 2007 and 2010, four percentage points higher than the state average. Also, from 2007 to 2010, achievement gaps were decreased by 11 percentage points between African-American and White students in high school reading. Additionally, 62% of African-American seniors participated in the SAT exam in 2010, ranking the district higher than the other 75 Broad Prize eligible urban school districts. The district has also received a

grant from the Wallace Foundation which will aid in the development of school leadership and measurement of student achievement (School District, 2013a).

Schools. The school district consists of 160 schools throughout the cities and towns of the serviced county. Of those, 89 are elementary schools with students in kindergarten through fifth grades; 39 are middle schools with students in sixth through eighth grades; 28 are high schools serving students in ninth through twelfth grades; and four are alternative schools serving a wide range of students. In the district's 160 schools, there are nine magnet programs housed in 37 different schools.

The average number of students in an elementary school in the district is 708. The average number of students in a middle school in the district is 1,002; and there are on average 1,251 students in a high school.

Students. In the district, there are currently 141,171 students in kindergarten through twelfth grades served by the school district. Table 1 describes the distribution of student population in those grades in the school district.

Table 1

Breakdown of Student Population

Grade Level	Count (N)	Percentage
Kindergarten-fifth grades	69,585	49%
Sixth-eighth grades	31,964	23%
Ninth-twelfth grades	39,622	28%
Total	141,171	

In the 2012-2013 school year, the school district was responsible for graduating 8,941 students. The student population is very diverse. Student demographics are wide-ranging, with the majority of the population characterized as Black/African American,

White, or Hispanic. Table 2 describes the student demographic breakdown within the district. Students represent 157 different countries, cultural, and ethnic backgrounds. There are 169 native languages spoken and approximately 14,830 students who are considered Limited English Proficient. Additionally, there are 76,232 students identified as economically disadvantaged in the district, representing 54% of the total student population (School District, 2013b).

Table 2

Student Demographics of District

Race	Count (N)	Percentage
Black/African American	59,291	42%
White	45,175	32%
Hispanic	25,411	18%
Asian	7,059	5%
Other	4,235	3%

Teachers. The school district is one of the largest employers in the county. Currently, there are 18,143 teachers, support staff, and administrators serving students across the school district. Of those, 9,221 are full-time teachers; 6,429 are support staff; and 995 are administrators and office staff. The district employs 9,180 certified teachers who average 10 1/2 years of experience. The district also encourages advanced learning as 3,547 have advanced degrees and 1,237 are National Board Certified (School District, 2013b). The majority of teachers in the district are classified as White or Black/African American. Table 3 shows the demographic distribution of the 9,221 full-time teachers in the district.

Table 3

Teacher Demographics of District

Race	Count (N)	Percentage
Black/African American	2,121	23%
White	6,547	71%
Hispanic	184	2%
Asian	92	1%
Other	277	3%
Total	9,221	

Collection and Analysis of Data

Data instrument. Data were collected from the state teacher evaluation instrument. This instrument consists of two major components. The first component is standards one through five which are comprised of observation ratings from school administration on elements of teacher effectiveness. The second component is standard six which is comprised of student growth data as measured by the EVAAS rating (DPI, 2013c).

Data for the educator evaluation system are captured annually in an online tool, mostly through observations conducted by administrators. The purpose of the educator evaluation system is to promote leadership, quality teaching, and student learning through a fair growth model. The evaluation process is based on gathering information from multiple data sources, employee artifacts, and other evidences to measure employee performance and effectiveness. One component of the system is that employees are responsible for setting performance goals as they strive to grow through professional development opportunities (DPI, 2013d).

Professional standards were developed for all employees of the state's public

school system. Professional organizations, staff members, and representatives from higher education were able to provide input as the standards were written. The standards include the skills, knowledge, and behaviors expected for each given role. Five standards were developed for teachers. On each of the standards, the teacher is given an overall rating on their summative review. Teachers receive one of five different ratings for each standard based on their observations and contributions to the school throughout the year. Teachers can receive ratings of Not Demonstrated, Developing, Proficient, Accomplished, and/or Distinguished. The state expectation is that every professional should obtain a rating of at least proficient on the standards (DPI, 2013d).

The educator evaluation system recently incorporated a sixth standard to capture teacher impact on student growth. The sixth standard rating is determined by data from EVAAS. The data are computed into a statistical number, or index, that can be used to compare teachers across grades, subjects, schools, and districts in the state. Based on those comparisons, teachers are given a rating of Does Not Meet Expectations, Meets Expectations, or Exceeds Expectations. For the purposes of this study, the value-added index was collected for analysis, not the comparison rating (DPI, 2013c). The value-added index is a continuous variable as scores exist along the continuum of scores from low to high (Creswell, 2008). On this standard, a rating of Does Not Meet Expected Growth consists of an index score of less than -2.0. A rating of Meets Expected Growth consists of an index score between -2.0 and +2.0. Finally, a rating of Exceeds Expected Growth consists of an index score of greater than +2.0 (DPI, 2013a).

Collection method. Quantitative data were collected on each fourth- through twelfth-grade teacher from the current state evaluation process. Teacher evaluation data were collected from existing summative evaluation reports from the 2012-2013 school

year. All teacher evaluation data collected by evaluators are housed in an online tool that stores and analyzes data. The data were collected and reported anonymously.

EVAAS data from 2012-2013 were collected from reports generated from the state value-added measurement for each teacher. Once assessments were conducted at the school level, assessments were scanned and reports were submitted to the state department of instruction. Once the state department reviewed the assessment data, they analyzed the initial data to verify reliability and validity of each assessment given. Any test that was deemed unreliable or invalid was removed from the data. The State Board of Education approved reports and sent the raw scores to SAS for the value-added scores to be determined. SAS also conducted a verification process to determine the validity and reliability of the assessments during their initial review of the data. If they deemed any of the assessments not valid or reliable during their statistical calculation of the EVAAS score, the results were not included when teacher index scores were computed. Once SAS, the independent vendor, analyzed the data and completed the reports, it reported the data back to the state board and department of instruction. The state released reports to districts, schools, and teachers and then populated standard six on evaluations with the data received. For this study, district reports were gathered for analysis (DPI, 2013d).

Data from the two instruments were compared to find the teachers who have summative evaluation data from the teacher evaluation instrument and EVAAS data from tested subjects. EVAAS data are available for fourth- through eighth-grade teachers who administer a state standardized end-of-grade math, reading, or science state assessment; high school teachers of courses with a state standardized Math I, English II, or biology assessment; or teachers who administer one of the teacher-created state common exams

or CTE postassessments. Those tests are given to students in the following courses: English Language Arts III, Social Studies, World History, Civics and Economics, Chemistry, U.S. History, English Language Arts III, Physics, Earth/Environmental Science, English Language Arts I, Physical Science, Business Law, Multimedia and Webpage Design, Principles of Business and Finance, Marketing, Interior Design I, Fashion Merchandising, Hospitality and Tourism, Sports & Entertainment Marketing II, Microsoft Word PowerPoint and Publisher, Medical Sciences I, Accounting I, Microsoft Excel and Access, Drafting-Architectural II, Drafting I, Scientific and Technical Visualization I, Biomedical Technology, Apparel I, Apparel II Enterprise, Early Childhood Education I, Culinary Arts and Hospitality I, Culinary Arts and Hospitality II, Foods I, Parenting and Child Development, Environmental and Natural Resources I, Horticulture I, and Horticulture II (DPI, 2013a).

Statistical method. The correlational study determined the relationship of teacher effectiveness as determined by teacher evaluation ratings and student growth as measured by EVAAS data. The quantitative study utilized data gathered from the reports and educator evaluation system reports. A quantitative correlation was deemed the most effective method for this research as it offered an objective approach and had the ability to identify statistically significant relationships or the strength of the association between the variables in the study. Traditionally, subjective dispositions arise when analyzing teacher evaluation data. However, the quantitative correlation approach provided the study with the statistical information to objectively highlight the relationship between the measures utilized to evaluate teachers (Creswell, 2008).

As the study includes two independent variables, and those two variables represent paired observations, the Spearman rank-order coefficient statistical test was

utilized to analyze the data. Spearman's correlation calculates a coefficient, r_s or ρ , which is a measure of the strength and direction of the association between one ordinal and one continuous variable (Creswell, 2008). The score on each standard of the teacher evaluation rubric is an ordinal value, whereas the EVAAS index score is a continuous variable. There also needs to be a monotonic relationship determined between the two variables in order to utilize Spearman's correlation. A monotonic relationship is a relationship that is defined by the value of one variable increasing as the other variable increases also or the value of one variable increasing and the other variable value decreasing. This assumption was checked by visually inspecting a scatterplot of the variable data. The Figure provides examples of monotonic and non-monotonic relationships (Lund Research, 2013).

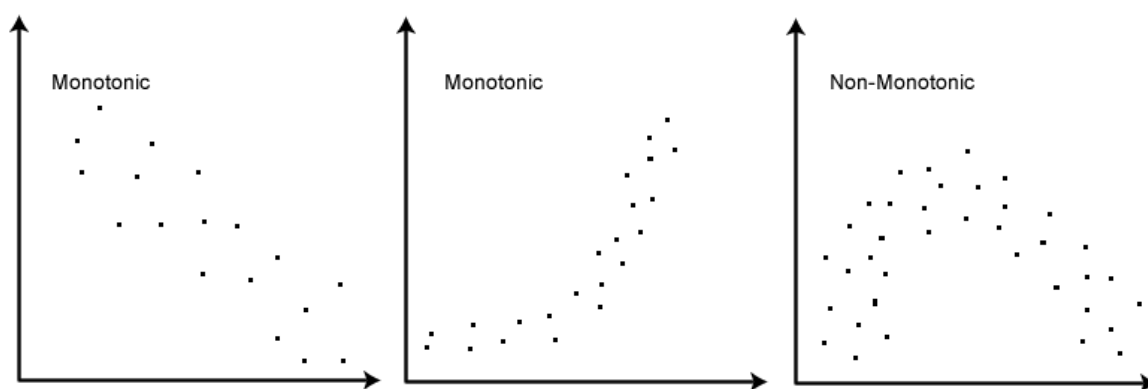


Figure. Monotonic and Non-Monotonic Relationships.

If the relationship was deemed to be monotonic, the researcher would conduct Spearman's correlation to determine whether a statistically significant relationship was present between the teacher observation instrument evaluations and the value-added scores for each teacher. Spearman's rank-order coefficient returned a value between

negative (-) 1.0 and positive (+) 1.0. Based on the value of r_s , the closer to +1.0 the findings yield, the greater the positive significant relationship between the two instruments. For example, if the value of r_s was greater than +0.50, a strong positive relationship existed between measures (Lund Research, 2013). Likewise, if the value of r was between +0.19 and -0.19, no relationship between instruments would be apparent (Cohen, 1988).

Once Spearman's correlation was calculated to describe the relationship between the ordinal and continuous variables, the researcher determined the statistical significance. The level of statistical significance is measured by p . When the statistical significance, p , is less than 0.05, it can be concluded that the correlation coefficient is statistically significantly different from zero (Lund Research, 2013).

Table 4

Evaluation Plan

Evaluation questions	Information required	Information source	Method of collecting information	Analysis procedures	Interpretation procedures and criteria
What is the relationship between a teacher's effectiveness rating on the state teacher evaluation instrument and the rating determined from value-added data?	Teacher evaluation data from state instrument and teacher value-added data	State evaluation instrument and state EVAAS data	Existing data – Reports will be collected for analysis	Correlation coefficient statistical test	Review of statistical data
What is the relationship between a teacher's score for leadership as defined by the state teacher evaluation instrument and the rating determined from value-added data?	Teacher evaluation data from state instrument and teacher value-added data	State evaluation instrument and state EVAAS data	Existing data – Reports will be collected for analysis	Correlation coefficient statistical test	Review of statistical data
What is the relationship between a teacher's score for establishment of a respectful classroom environment for a diverse population as defined by the state teacher evaluation instrument and the rating determined from value-added data?	Teacher evaluation data from state instrument and teacher value-added data	State evaluation instrument and state EVAAS data	Existing data – Reports will be collected for analysis	Correlation coefficient statistical test	Review of statistical data
What is the relationship between a teacher's score for knowledge of the content they teach as defined by the state teacher evaluation instrument and the rating determined from value-added data?	Teacher evaluation data from state instrument and teacher value-added data	State evaluation instrument and state EVAAS data	Existing data – Reports will be collected for analysis	Correlation coefficient statistical test	Review of statistical data

(continued)

Evaluation questions	Information required	Information source	Method of collecting information	Analysis procedures	Interpretation procedures and criteria
What is the relationship between a teacher's score for ability to facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data?	Teacher evaluation data from state instrument and teacher value-added data	State evaluation instrument and state EVAAS data	Existing data – Reports will be collected for analysis	Correlation coefficient statistical test	Review of statistical data
What is the relationship between a teacher's score for reflection on his/her practice as defined by the state teacher evaluation instrument and the rating determined from value-added data?	Teacher evaluation data from state instrument and teacher value-added data	State evaluation instrument and state EVAAS data	Existing data – Reports will be collected for analysis	Correlation coefficient statistical test	Review of statistical data

Limitations

There were limitations that impacted the methodology of this study. There was a general decrease in the overall sample size because not all teachers had both data points available. Some teachers in the district did not have usable teacher evaluation data from the educator evaluation system rubric. There were a variety of circumstances that eliminated summative data points from the sample such as misadministered evaluations, incorrect processing or filing of evaluations, or mid-year separations that result in an incomplete summative evaluation. Similarly, there are two groups of teachers in the district who did not currently have EVAAS data: teachers who do not administer district or state common final exams and teachers of kindergarten through second-grade students. Any teacher who did not have both educator effectiveness data and EVAAS data could not be included in the correlation. Of the 9,221 full-time teachers in the district, 3,338

had summative evaluation data and EVAAS data. Therefore, after excluding the missing data, the sample size was large enough to determine a correlation for the purposes of this study.

In an ideal implementation of the educator evaluation system, all evaluations are completed objectively with no bias. However, because of the subjective nature of the rubric and the sheer number of evaluators across the district, one limitation exists with the rater reliability. Although all evaluators participate in training and orientation sessions provided by the district, there is still a variance in rubric scores.

Similarly, for the purposes of this study, the teacher composite EVAAS score was used to assess the relationship between student performance and teacher effectiveness as determined by teacher evaluations. All of the courses, sections, and different content areas taught by a teacher were included in the composite score. For example, the composite score of an elementary mathematics teacher may include one math class of 25 students. However, the composite score for a high school mathematics teacher may include two geometry courses and three algebra courses, representing over 100 students. Due to the nature of the data, it was not possible to disaggregate those scores any further. As a result of the sample size of the data, however, the integrity of the findings was not compromised.

Delimitations

A delimitation of this study is the exclusion of qualitative data. Based on the methodology of this study, the educator effectiveness data and EVAAS data are existing sources from previous school years. As a result, it was decided not to include surveys or focus-group dialog with current principals because those data were not available during the time of initial collection of the data. Because of climate and culture changes with

staff retention and turnover, the addition of qualitative data would not contribute to the accuracy of the study.

Summary

This study aimed to find the correlation between the two measures of teacher effectiveness currently being used because of state requirements. The state utilizes a teacher evaluation instrument that rates teachers on five standards as a result of principal observations and a sixth standard of a value-added model score based on student growth data (DPI, 2013c). The study determined the relationship between the two measures in the single district. As research explains, the teacher has the greatest impact on student achievement (McCaffrey et al., 2003); and this study strived to better understand how they are evaluated.

Chapter 4: Results

Introduction

This study determined the relationship between ratings on the state teacher evaluation instrument and value-added index scores among teachers of tested subjects in Grades 4-12 in a large, urban school district in a southeastern state. Administrator observations informed the ratings given on standards one through five on the educator evaluation system. A new value-added initiative was implemented using data from a private vendor to populate standard six on the educator evaluation system rubric. Through this research study, the researcher aimed to identify the current state of teacher evaluation and value-added ratings in the district, as well as determine the extent of the relationship between these two measures. In this chapter, the research questions and findings for each relationship are discussed.

Research Questions

1. What is the relationship between a teacher's effectiveness rating on the state teacher evaluation instrument and the rating determined from value-added data?
2. What is the relationship between a teacher's score for leadership as defined by the state teacher evaluation instrument and the rating determined from value-added data?
3. What is the relationship between a teacher's score for establishment of a respectful environment for a diverse population of students as defined by the state teacher evaluation instrument and the rating determined from value-added data?
4. What is the relationship between a teacher's score for knowledge of the

content they teach as defined by the state teacher evaluation instrument and the rating determined from value-added data?

5. What is the relationship between a teacher's score for ability to facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data?
6. What is the relationship between a teacher's score for reflection on his/her practice as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Data Collection Process

Quantitative data were collected prior to and independent of this study on each teacher, fourth through twelfth grades, during the 2012-2013 school year as a result of the current state evaluation process. Summative evaluation reports, including ratings on standards one through five, were retrieved from the online tool that stores and analyzes data for district purposes. EVAAS data from 2012-2013, as represented in standard six, were also collected from reports generated from the state value-added measurement for each teacher. Assessments were conducted at the school level, then scanned and submitted to the State Department of Instruction. The state department reviewed and approved the assessment data, verifying reliability and validity; and sent raw scores to SAS, the private vendor, for the value-added scores to be determined. SAS also determined the validity and reliability of the assessments during their review and ran all verified data through their model to find the statistical calculation, or value-added score. The data were reported back to the State Board and Department of Instruction then subsequently released to districts, schools, and teachers (DPI, 2013d). District reports were then retrieved for the purposes of this study. Data from the two instruments were

merged, isolating a sample of only teachers who had both summative evaluation data from the teacher evaluation instrument and EVAAS data from tested subjects.

Sample Demographics

Of the 9,221 full-time teachers in the district, there were 3,338 teachers who had both summative evaluation data and EVAAS data for the 2012-2013 school year. The sample is representative of the overall population of the district, with the majority of teachers being White females. Based on the demographics of teachers in the entire district, 71% were White; 23% Black/African American; 2% Hispanic; 1% Asian; and 3% other. Similarly, in the sample included in the data analysis, 69% were White; 26% Black/African American; 1% Hispanic; 1% Asian; and 3% other. Of the 3,338 teachers in the study, 62% are experienced teachers and have received career status in the state; while 38% are probationary, or beginning, teachers with only 1-3 years of experience. Table 5 shows the demographic distribution of the 3,338 full-time teachers in the district who were included in this study.

Table 5

Sample Teacher Demographics

	Count (N)	Percentage
Gender		
Male	777	23%
Female	2,561	77%
Race		
White	2,311	69%
Black/African American	856	26%
Hispanic	44	1%
Asian	32	1%
Other	95	3%
Experience		
Probationary	1,284	38%
Career	2,054	62%
School Level		
Elementary School	1,085	33%
Middle School	1,135	34%
High School	1,111	33%
Other	7	<1%

Additionally, as the data are required to show paired observations between the teacher effectiveness data and the value-added data, Table 6 shows the sample size available for analysis for each standard on the evaluation tool. The sample sizes noted in the table were used for analysis to answer the research questions.

Table 6

Teacher Sample Size

Evaluation Standard	Sample (N)
Standard I	3,330
Standard II	2,126
Standard III	2,108
Standard IV	3,329
Standard V	2,110
EVAAS/Standard VI	3,338

As Table 6 shows, the sample sizes vary for each standard analyzed. Standards one and four were captured for over 99% of the teachers who had reported EVAAS data populating standard six. However, standards two, three, and five had much less representation from the whole sample. In the state, administrators are only required to evaluate career, or tenured, teachers on standards one and four every year. On teacher's renewal year, which occurs every fifth year of teaching, the process requires the evaluation to include a rating for each standard (DPI, 2012e). As a result, career teachers who are not in their renewal year did not have data reported for standards two, three, and five. Standards two, three, and five include data from probationary, or beginning, teachers and career teachers in their renewal year.

Monotonic Relationship

The Spearman rank-order coefficient was conducted to analyze the data as the study included two independent variables and those two variables represent paired observations (Creswell, 2008). The data were represented on a scatterplot to determine if there was a monotonic relationship between the two variables. A monotonic relationship is a relationship that is defined by the value of one variable increasing as the other

variable also increases or the value of one variable increasing and the other variable value decreasing. The scatterplot provided evidence of an existing monotonic relationship.

This assumption was checked by a visual inspection of all of the variables (see Appendix B).

Findings

The degree of the Spearman coefficient determines the strength of the correlation. Although assigning strength of association slightly varies based on the study, Spearman coefficients tend to be smaller than other correlations such as Pearson. To define the relationship between each standard and the EVAAS data, the classifications identified in Table 7 were used based on the value returned from Spearman's rank-order coefficient (Lund Research, 2013).

Table 7

Determining the Strength of the Relationship

Value of r_s	Strength of Relationship
+0.70 or higher	Very strong positive relationship
+0.40 to +0.69	Strong positive relationship
+0.30 to +0.39	Moderate positive relationship
+0.20 to +0.29	Weak positive relationship
+0.01 to +0.19	Negligible positive relationship
-0.01 to -0.19	Negligible negative relationship
-0.20 to -0.29	Weak negative relationship
-0.30 to -0.39	Moderate negative relationship
-0.40 to -0.69	Strong negative relationship

Research Question 1. What is the relationship between a teacher's effectiveness rating on the state teacher evaluation instrument and the rating determined from value-added data?

Spearman's rank-order coefficient was calculated to find the relationship between each standard on the teacher evaluation rubric and the EVAAS score as defined in standard six of the teacher evaluation. Based on the findings in Table 8, the relationships are all positive. However, it can be stated that overall there is little to no relationship found between the EVAAS data and any of the evaluated standards on the rubric. As the standards are rated independent of each other based on the qualifying elements previously discussed in Chapter 3, there is not an overall mean score generated based on standards one through five to analyze separately with the EVAAS data. However, the researcher found that based on the results from the following research questions, although the coefficients found are positive, there is a negligible to weak relationship between a teacher's effectiveness rating on the state teacher evaluation instrument and the rating determined from value-added data. The relationship between EVAAS and each standard are specifically addressed in Research Questions 2-6.

Research Question 2. What is the relationship between a teacher's score for leadership as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 2, the relationship between a teacher's leadership as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant weak, positive correlation. As the researcher hypothesized, the data did not show a strong correlation with the value-added scores for standard one. Spearman's coefficient returned a value of $r_s=0.21$ with $df=3,328$ and $p=0.00$, making the coefficient statistically significant (Lund Research, 2013). See Table 8.

Research Question 3. What is the relationship between a teacher's score for

establishment of a respectful environment for a diverse population of students as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 3, the relationship between a teacher's establishment of a respectful environment for a diverse population of students as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant weak, positive correlation. As the researcher hypothesized, the data did not show a strong correlation with the value-added scores for standard two. Spearman's coefficient returned a value of $r_s=0.26$ with $df=2,124$ and $p=0.00$, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 8.

Research Question 4. What is the relationship between a teacher's score for knowledge of the content they teach as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 4, the relationship between a teacher's knowledge of the content they teach as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant weak, positive correlation. As the researcher hypothesized, the data did not show a strong correlation with the value-added scores for standard three. Spearman's coefficient returned a value of $r_s=0.21$ with $df=2,106$ and $p=0.00$, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 8.

Research Question 5. What is the relationship between a teacher's score for ability to facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 5, the relationship between a teacher's ability to facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant weak, positive correlation. As the researcher hypothesized, the data did not show a strong correlation with the value-added scores for standard four. Spearman's coefficient returned a value of $r_s=0.25$ with $df=3,327$ and $p=0.00$, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 8.

Research Question 6. What is the relationship between a teacher's score for reflection on his/ her practice as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 6, the relationship between a teacher's ability to facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant negligible, positive correlation. As the researcher hypothesized, the data did not show a strong correlation with the value-added scores for standard five. Spearman's coefficient returned a value of $r_s=0.19$ with $df=2,110$ and $p=0.00$, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 8.

Table 8

Spearman's Coefficient

	Standard I	Standard II	Standard III	Standard IV	Standard V	EVAAS
Standard I	1.00					
Standard II	0.66 0.00	1.00				
Standard III	0.59 0.00	0.60 0.00	1.00			
Standard IV	0.70 0.00	0.68 0.00	0.70 0.00	1.00		
Standard V	0.65 0.00	0.60 0.00	0.59 0.00	0.65 0.00	1.00	
EVAAS	0.21 0.00	0.26 0.00	0.21 0.00	0.25 0.00	0.19 0.00	1.00

Additional Correlations

Based on the data collected from the district and the initial correlation results discussed in Chapter 4, the researcher conducted additional correlations. The researcher calculated Spearman's coefficient based on school level, gender, ethnicity, and career status.

As aforementioned, there were 3,338 teachers of the 9,221 full-time teachers in the district who have both summative evaluation data and EVAAS data for the 2012-2013 school year. That sample is representative of the overall population of the district.

School level. As the data are required to show paired observations between the teacher effectiveness data and the value-added data, Table 9 shows the sample size available for analysis for each standard on the evaluation tool based on teacher school level. The sample sizes noted in the table were used for analysis to find the additional correlations.

Table 9

School-Level Sample Sizes

	Sample (N)
Elementary	
Standard I	1,083
Standard II	644
Standard III	637
Standard IV	1,083
Standard V	638
EVAAS	1,085
Middle School	
Standard I	1,134
Standard II	749
Standard III	742
Standard IV	1,134
Standard V	742
EVAAS	1,135
High School	
Standard I	1,106
Standard II	726
Standard III	722
Standard IV	1,105
Standard V	723
EVAAS	1,111

Analysis of the school-level correlations aimed to find whether there was a stronger relationship between observation data for each standard and EVAAS data at any particular level. For elementary school teachers, Spearman's coefficient provides evidence of a statistically significant weak, positive correlation for standards one through four, mirroring the data discussed in Chapter 4. Similarly, the standard five coefficient returned a value of $r_s=0.17$ with $df=636$, showing a negligible relationship. Spearman's also returned $p<0.01$ for each standard, making the correlation coefficient statistically

significant (Lund Research, 2013). As aforementioned, the relationships found between the standards based on principal observation and EVAAS data for elementary school teachers showed relationships that were representative of the overall findings. One difference found among the elementary teacher sample was the strongest relationship existed with standard four as $r_s=0.28$ with $df=1,081$. See Table 10.

However, the findings for middle school teachers and high school teachers did not provide the same results. Unlike the overall findings and the results from elementary school teachers, the middle school teacher coefficient for standard one, teachers demonstrate leadership, returned a value of $r_s=0.19$ with $df=1,132$, showing a negligible relationship between the standard and EVAAS scores. The remaining four standards, standards two through five, returned positive, weak correlations with the value of r_s ranging from 0.21 to 0.28. Previously with the overall sample, standard five was determined to have a negligible relationship. The standards for high school teachers all returned correlations values ranging from $r_s=0.21$ to $r_s=0.24$, showing all five standards have a positive, weak correlation with the EVAAS data. See Table 10.

Table 10

School-Level Correlations

	Standard I	Standard II	Standard III	Standard IV	Standard V
Elementary Teachers EVAAS	0.20 0.00	0.25 0.00	0.20 0.00	0.28 0.00	0.17 0.00
Middle School Teachers EVAAS	0.19 0.00	0.28 0.00	0.26 0.00	0.25 0.00	0.21 0.00
High School Teachers EVAAS	0.21 0.00	0.24 0.00	0.21 0.00	0.22 0.00	0.23 0.00

Gender. As the data are required to show paired observations between the teacher effectiveness data and the value-added data, Table 11 shows the sample size available for analysis for each standard on the evaluation tool based on teacher gender. The sample sizes noted in the table were used for analysis to find the additional correlations.

Table 11

Gender Sample Sizes

Evaluation Standard	Sample (N)
Male Teachers	
Standard I	771
Standard II	503
Standard III	501
Standard IV	771
Standard V	501
EVAAS	777
Female Teachers	
Standard I	2,559
Standard II	1,623
Standard III	1,607
Standard IV	2,558
Standard V	1,609
EVAAS	2,561

Analysis of the gender correlations aimed to find whether there was a stronger relationship between observation data for each standard and EVAAS data for males or females. For male teachers, Spearman's coefficient provides evidence of a statistically significant weak, positive correlation for standards one through four, mirroring the data discussed in Chapter 4. Similarly, the standard five coefficient returned a value of $r_s=0.16$ with $df=499$, showing a negligible relationship. Spearman's also returned $p=<0.01$ for each standard, making the correlation coefficient statistically significant (Lund Research, 2013). As aforementioned, the relationships found between the standards based on principal observation and EVAAS data for male teachers showed relationships that were representative of the overall findings. See Table 12.

However, the findings for female teachers yielded differing results. Unlike the

overall findings and the results from male teachers, the female teacher coefficient for standard one, teachers demonstrate leadership, returned a value of $r_s=0.18$ with $df=2,557$; showing a negligible relationship between the standard and EVAAS scores. Standards two through four returned positive, weak correlations with the value of r_s ranging from 0.21 to 0.25. Previously with the overall sample, standard five was determined to have a negligible relationship. Similarly, standard five among female teachers returned a value of $r_s=0.18$ with $df=1,609$, also showing a negligible relationship. As was the case with the male results, Spearman's returned $p=<0.01$ for each standard, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 12.

Table 12

Gender Correlations

	Standard I	Standard II	Standard III	Standard IV	Standard V
Male Teachers EVAAS	0.23 0.00	0.26 0.00	0.21 0.00	0.25 0.00	0.16 <0.01
Female Teachers EVAAS	0.18 0.00	0.25 0.00	0.21 0.00	0.24 0.00	0.19 0.00

Ethnicity. As the data are required to show paired observations between the teacher effectiveness data and the value-added data, Table 13 shows the sample size available for analysis for each standard on the evaluation tool based on teacher ethnicity. The sample sizes noted in the table were used for analysis to find the additional correlations.

Table 13

Ethnicity Sample Sizes

Evaluation Standard	Sample (N)
White Teachers	
Standard I	2,306
Standard II	1,456
Standard III	1,443
Standard IV	2,306
Standard V	1,443
EVAAS	2,311
Black/African-American Teachers	
Standard I	854
Standard II	551
Standard III	546
Standard IV	853
Standard V	547
EVAAS	856
Hispanic Teachers	
Standard I	44
Standard II	37
Standard III	37
Standard IV	44
Standard V	37
EVAAS	44
Asian Teachers	
Standard I	31
Standard II	28
Standard III	28
Standard IV	31
Standard V	28
EVAAS	32
Other Teachers	
Standard I	95
Standard II	54
Standard III	54
Standard IV	95
Standard V	55
EVAAS	95

Analysis of the ethnicity correlations aimed to find whether there was a stronger

relationship between observation data for each standard and EVAAS data based on ethnicity. For White teachers, Spearman's coefficient provides evidence of a statistically significant weak, positive correlation for standards one through four, mirroring the data discussed in Chapter 4. Similarly, the standard five coefficient returned a value of $r_s=0.18$ with $df=1,441$, showing a negligible relationship. Spearman's also returned $p<0.01$ for each standard, making the correlation coefficient statistically significant (Lund Research, 2013). As aforementioned, the relationships found between the standards based on principal observation and EVAAS data for White teachers showed relationships that were representative of the overall findings. See Table 14.

However, the findings for Black/African-American teachers yielded different results. Unlike the overall findings and the results from White teachers, the Black/African-American teacher coefficient for standard one, teachers demonstrate leadership, returned a value of $r_s=0.16$ with $df=852$; showing a negligible relationship between the standard and EVAAS scores. Standards two through five returned positive, weak correlations with the value of r_s ranging from 0.20 to 0.28. Previously with the overall sample, standard five was determined to have a negligible relationship but has a weak, positive relationship among Black/African-American teachers. Spearman's returned $p<0.01$ for each standard, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 14.

Because of the smaller sample sizes, not all of the Hispanic teacher correlations returned a statistically significant coefficient as $p>0.05$. However, a positive, moderate relationship was found among Hispanic teachers for standards three and four, as $r_s=0.39$ and 0.31, respectively. For standard three, $df=35$ and $p=0.02$; and for standard four, $df=42$ and $p=0.04$.

The findings for Asian teachers yielded different results than any of the previous analyses. Pearson's coefficient returned a positive, weak correlation for standard one as $r_s=0.28$ with $df=29$. However, for standard two, which has previously returned one of the highest correlation coefficients, returned a value of $r_s=0.03$ with $df=26$; showing a negligible relationship between standard two and EVAAS data. However, for standards three through five, the correlations returned were the strongest found. For standard three, $r_s=0.34$ with $df=26$; for standard four, $r_s=0.47$ with $df=29$; and for standard five, $r_s=0.36$ with $df=26$. The values all represent a positive, moderate relationship. However, Spearman's did not return a statistically significant value for standards one, two, and three as $p>0.05$. Standards four and five returned a value of $p<0.05$, making those correlation coefficients statistically significant (Lund Research, 2013). See Table 14.

Once again, because of the smaller sample sizes for teachers of other ethnicities, not all of the correlation coefficients returned were statistically significant. Standard one returned $p=0.04$, and standard four returned $p=0.00$; making each statistically different from zero. Both correlations were also found to be positive with standard one showing a weak correlation as $r_s=0.22$ with $df=93$, and standard four showing a moderate correlation as $r_s=0.33$ with $df=93$. See Table 14.

Table 14

Ethnicity Correlations

	Standard I	Standard II	Standard III	Standard IV	Standard V
White Teachers					
EVAAS	0.22 0.00	0.25 0.00	0.20 0.00	0.26 0.00	0.18 0.00
Black/African-American Teachers					
EVAAS	0.16 0.00	0.28 0.00	0.23 0.00	0.20 0.00	0.24 0.00
Hispanic Teachers					
EVAAS	0.02 0.91	0.31 0.06	0.39 0.02	0.31 0.04	0.04 0.81
Asian Teachers					
EVAAS	0.28 0.13	0.03 0.87	0.34 0.07	0.47 <0.01	0.36 0.05
Other Teachers					
EVAAS	0.22 0.04	0.21 0.13	0.14 0.30	0.33 <0.01	0.13 0.34

Career status. As the data are required to show paired observations between the teacher effectiveness data and the value-added data, Table 15 shows the sample size available for analysis for each standard on the evaluation tool based on teacher career status. The sample sizes noted in the table were used for analysis to find the additional correlations.

Table 15

Career Status Sample Sizes

Evaluation Standard	Sample (N)
Probationary Teachers	
Standard I	1,279
Standard II	1,279
Standard III	1,278
Standard IV	1,278
Standard V	1,278
EVAAS	1,284
Career Teacher	
Standard I	2,051
Standard II	847
Standard III	830
Standard IV	2,051
Standard V	832
EVAAS	2,054

Analysis of the career status correlations aimed to find whether there was a stronger relationship between observation data for each standard and EVAAS data based on career status. For probationary teachers, Spearman's coefficient provides evidence of a statistically significant weak, positive correlation for standards one, two, four, and five, as the value of r_s ranged from 0.20 to 0.27. The standard three coefficient returned a value of $r_s=0.18$ with $df=1,276$, showing a negligible relationship. Spearman's also returned $p<0.01$ for each standard, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 16.

The findings for career teachers yielded different results. Unlike the overall findings and the results from probationary teachers, the career teacher coefficient for standard one, teachers demonstrate leadership, returned a value of $r_s=0.19$ with $df=2,049$;

showing a negligible relationship between the standard and EVAAS scores. Standards two through five all showed weak, positive relationships among career teachers. For standard two, $r_s=0.27$ with $df=845$; for standard three, $r_s=0.25$ with $df=828$; for standard four, $r_s=0.25$ with $df=2,049$; and for standard five, $r_s=0.20$ with $df=830$. As was the case with the probationary results, Spearman's returned $p=<0.01$ for each standard, making the correlation coefficient statistically significant (Lund Research, 2013). See Table 16.

Table 16

Career Status Correlations

	Standard I	Standard II	Standard III	Standard IV	Standard V
Probationary Teachers EVAAS	0.22 0.00	0.27 0.00	0.18 0.00	0.26 0.00	0.20 0.00
Career Teachers EVAAS	0.19 0.00	0.27 0.00	0.25 0.00	0.25 0.00	0.20 0.00

Summary

The results of Spearman's rank-order coefficient proved that for standards one through four on the state teacher evaluation instrument there were positive, weak relationships found with the EVAAS, or standard six, data. The findings on standards one through four ranged from $r_s=0.26$ to $r_s=0.21$. However, when calculated for standard five, Spearman's correlation indicated there was a negligible relationship found with the EVAAS data as $r_s=0.19$. However, all of the calculations were found to be statistically significant as $p=<0.01$ for all standards.

Chapter 5: Discussion

Introduction

This correlational study was conducted to determine the relationship between two measures of teacher effectiveness in a large, urban school district in a southeastern state in the United States. Teachers have been evaluated using a teacher evaluation instrument that rates teachers on five standards based on principal observations. As part of the instrument, the state recently implemented the use of a value-added model that measures a teacher's impact on student learning based on student achievement on standardized tests which populates standard six on that rubric (DPI, 2013c). The value-added data and the state teacher evaluation instrument data were analyzed among fourth- through twelfth-grade teachers who taught subjects that evaluated students using an end-of-grade or end-of-course state-standardized assessment.

Summary of Findings

Once the relationship was deemed to be monotonic, the researcher conducted Spearman's correlation to determine whether a statistically significant relationship was present between the teacher observation instrument evaluations and the value-added scores for each teacher. Spearman's rank-order coefficient returns a value between negative (-) 1.0 and positive (+) 1.0. Based on the value of r_s , the closer to +1.0 the findings yield, the greater the positive significant relationship between the two instruments (Lund Research, 2013). For standard one, the value of r_s was +0.21, indicating a positive, weak relationship existed between measures. For standard two, the value of r_s was +0.26, indicating a positive, weak relationship existed between measures. For standard three, the value of r_s was +0.21, indicating a positive, weak relationship existed between measures. For standard four, the value of r_s was +0.25, indicating a

positive, weak relationship existed between measures. For standard five, the value of r_s was +0.19, indicating no significant relationship between instruments was apparent.

Once Spearman's correlation was calculated to describe the relationship between the ordinal and continuous variables, the researcher also determined the statistical significance measured by p . When the statistical significance, p , is less than 0.05, it can be concluded that the correlation coefficient is statistically significantly different from zero (Lund Research, 2013). For all five standards, $p < 0.01$, showing the study was statistically significant.

As a result of the Spearman's calculation, the researcher determined that the state teacher evaluation rubric, based on principal evaluation, does not identify the teachers with the highest value-added scores. In theory, the highest performing teachers should receive the highest ratings on both instruments. However, a lack of a strong positive correlation suggests there is not a significant relationship between the tools. Based on the findings, the researcher concluded that both of the measures, the teacher evaluation rubric and EVAAS data, still do not accurately identify and appropriately evaluate effective teaching.

Based on the findings, the researcher theorized three causes for the negligible and weak correlations found between the standards and EVAAS scores. First, the negligible and weak correlations could be caused by a lack of validity of the teacher evaluation instrument. Second, the lack of reliability of the evaluator ratings could cause the weaker correlations. Lastly, there could be little to no connection between the content of the standards and teacher effectiveness.

The researcher reviewed the research discussed in Chapter 2 in order to find which of the three aforementioned causes led to the negligible and weak correlations

found in the study. The research supported the use of an evaluation instrument in the district based on the findings from studies that found validity in the Danielson model, as did the 2011 study in Chicago and the 2006 study in Cincinnati, Los Angeles, Reno, and Coventry (Heneman et al., 2006).

Further studies, such as the 2013 MET study and the 2009 TNTP, published findings that illustrate the connection between the content of the teacher evaluation standards and teacher effectiveness. Additionally, WestEd (2012) conducted a review of the educator evaluation system that verified the validity based on six findings around student learning expectations, opportunities for learning, accurate and reliable measure of student achievement, meaningful interpretations of student growth, teachers' instructional effectiveness, and the inclusivity of the comprehensive system. The validity study provides evidence to support the use of the educator evaluation system in the state, which led the researcher to further research the implementation of the measure as it pertains to reliability of the evaluator ratings.

One emerging theme from the research highlights a weakness in administrator knowledge and training around measuring teacher effectiveness. The RAND, TNTP, and MET studies all found that measures of teacher effectiveness were valid and reliable when inter-rater reliability was increased through rigorous evaluator training. As a result, the researcher focused discussions on the improvement of evaluator implementation of the instrument. The following discussion considers each question specifically and then provides suggestions about improving the implementation process through training.

Research Questions

Research Question 1. What is the relationship between a teacher's effectiveness rating on the state teacher evaluation instrument and the rating determined from value-

added data?

Spearman's rank-order coefficient was calculated to determine the relationship between each standard on the teacher evaluation rubric and the EVAAS score as defined in standard six of the teacher evaluation. Based on the findings, the relationships are all positive. However, it can be stated that overall there is little to no relationship found between the EVAAS data and any of the evaluated standards on the rubric.

These findings are significant, as principals are rating teachers based on their understanding of effective teaching. If principals are appropriately identifying their most effective teachers, they should ultimately be able to predict which teachers' EVAAS scores will be the highest, as the two should be correlated. The following analyses elaborate on the findings for each standard.

Research Question 2. What is the relationship between a teacher's score for leadership as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 2, the relationship between a teacher's leadership as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant weak, positive correlation.

As thoroughly discussed in Chapter 2, standard one is a critical aspect of the evaluation as it assesses the participation of teachers in actively leading the school in collaboration with the administration (DPI, 2013d). Teachers are evaluated on their ability as classroom leaders to be responsible for the learning and progress of all students to ensure they are contributing to and imparting the vision to students that they graduate from high school and are ready to pursue postsecondary education and compete in a

global workforce. Part of that process includes the expectations that teachers are utilizing a variety of data sources to aid in the setting and development of goals that meet the needs of each student. As the year progresses, teachers are responsible for student learning as they evaluate student progress towards those goals and make adjustments as necessary (DPI, 2012e).

Danielson (1996) outlined these characteristics of teacher leadership as an essential responsibility of effective teachers within each domain of her framework for teaching as these practices directly impact student performance. As a result, the ratings given for standard one should correlate with EVAAS scores, as the content has a direct correlation with teacher effectiveness. As this study showed the two measures had a positive, weak relationship, there is cause to review the district indicators that guide administrators in providing the appropriate ratings. The elements of the standard should be closely analyzed to better identify what quality leadership looks like in the classroom and within a school. With continued improvement in training and resources, administrators should be equipped to identify the teachers who are leading their classrooms, schools, and the profession well to appropriately categorize the most effective teachers.

Research Question 3. What is the relationship between a teacher's score for establishment of a respectful environment for a diverse population of students as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 3, the relationship between a teacher's establishment of a respectful environment for a diverse population of students as defined by the state teacher evaluation instrument and the rating determined from value-added

data, provides evidence of a statistically significant weak, positive correlation.

The second standard in the teacher evaluation instrument assesses the classroom environment created by the teacher and whether classrooms are positive and nurturing in order to facilitate the development of a healthy teacher-student relationship. Standard two assesses the teacher responsibility for ensuring the learning environment is inviting, respectful, supportive, inclusive, and flexible. The standard explains the necessity to select materials and organize lessons that include contributions from a variety of cultures in order to embrace diversity. In doing so, teachers must recognize the influence of culture on a student's development, personality, and performance as well as understand the effects of race, ethnicity, gender, religion, and other cultural influences. The various elements captured within the standard consider that understanding, outlining the responsibility to consider those perspectives and dispositions when planning instruction (DPI, 2012e). Teachers are also held accountable for their treatment of students within standard two. Teachers are required to hold students to high expectations and adapt their practices to promote learning for all students, including those with special needs. Teachers are to carefully collaborate with colleagues to design instructional opportunities around the needs of their students that utilize best practices (DPI, 2012c).

The highest correlation existed between EVAAS and standard two, but it was still a weak correlation as $r_s=0.26$. As standard two identifies elements of quality teaching as it pertains to student needs and instruction, it should be strongly correlated to the EVAAS data. Although standard two aims to evaluate teachers on the aforementioned characteristics, the correlation results indicate that the instrument is not a valid measure; there is no connection between the classroom environment and teacher effectiveness; or inter-rater reliability needs improvement. The WestEd (2012) study of the educator

evaluation system supported the development of the instrument and determined the rubric was an accurate and reliable measure of effectiveness. Similarly, the work of Danielson (1996) and the research conducted negate the likelihood of the second probable cause (Heneman et al., 2006).

As a result, the researcher suggests the weak correlations are caused by a lack of knowledge and training for the evaluator. The district indicators and elements should be reviewed within the district to better align evaluations with performance of effective teachers. A revision would provide principals with solid, observable characteristics; improving inter-rater reliability. Therefore, principals should be provided more guidance in observing teachers to ensure those rated at an accomplished level are utilizing those observed skills and best practices to increase student achievement.

Research Question 4. What is the relationship between a teacher's score for knowledge of the content they teach as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 4, the relationship between a teacher's knowledge of the content they teach as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant weak, positive correlation.

Standard three evaluates teachers on their content knowledge and instructional delivery. Elements provide details about the expectations to create engaging, relevant lessons that connect content to students' lives, integrate multiple disciplines, and include teaching 21st century content. The state provides a specific instructional guide, or course of study, for teachers to align lessons and activities with learning standards. Standard three assesses understanding of those standards and alignment of instruction (DPI,

2012e). Standard three holds teachers accountable for facilitating the learning of core content, 21st century content, and the development of higher-level thinking and 21st century skills (DPI, 2012c).

As standard three directly assesses the instructional content and delivery, in theory, it should have one of the strongest correlations with the value-added index score; however, Spearman's correlation showed that it had one of the weakest correlations with the EVAAS data. The lack of a strong relationship between this standard and the performance data would suggest that the evaluation tool be further analyzed to understand the causality. The research continues to support the supposition that administrators are not identifying the highest performing teachers using the rubric in its current form based on the state processes. The work around evaluating teacher effectiveness clearly identifies content pedagogy as a key aspect of teacher effectiveness (Danielson, 1996). As a result, the qualitative data captured through principal observation is not invaluable. However, evaluators should more accurately predict which teachers are showing the most growth with their students over the course of the school year as they observe the content being taught in classrooms. As the district continues working towards improvement in the evaluation process, evaluators need more specific training and resources around content and observing standards alignment of instruction. If teachers are effectively teaching the content, student achievement increases. Principals should appropriately capture that exchange through observations.

Research Question 5. What is the relationship between a teacher's score for ability to facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data?

Analysis of Research Question 5, the relationship between a teacher's ability to

facilitate learning for their students as defined by the state teacher evaluation instrument and the rating determined from value-added data, provides evidence of a statistically significant weak, positive correlation.

Standard four of the teacher evaluation rubric outlines the teacher's role as the facilitator of learning in the classroom. Within that responsibility, teachers are expected to allow students to create their own learning experiences through their planning of innovative lessons. Students should be provided opportunities to collaborate and communicate with their peers.

The elements outline expectations that teachers understand the intellectual, physical, social, and emotion development levels of their students so they are able to differentiate learning for their students. The standard explains the need for teachers to utilize a variety of instructional methods to plan instructional lessons and activities based on the needs of the students. Teachers are expected to integrate and utilize technology to maximize student learning and to promote critical thinking, problem solving, communication, innovation, collaboration while teaching appropriate usage and the ability to identify credible sources (DPI, 2012e).

Teachers are responsible for building higher-order thinking skills such as asking questions, drawing conclusions, making complex choices, and synthesizing information, in addition to teaching the skills necessary for students to be globally competitive. Additionally, teachers are expected to provide opportunities for students to collaborate with their peers, assume leadership roles, improve communication skills, and interact with people from different cultures and with different strengths (DPI, 2012e).

As standard four focuses on facilitating learning in the classroom, there should be a strong correlation between the learning standard and the value-added index score as a

result of student performance. If teachers are rated highly on standard four, they have been documented as teaching at a high level and exposing their students to best learning practices. However, the correlation for standard four and EVAAS was found to be from $r_s=0.25$, representing only a weak relationship. Because the findings did not result in a strong correlation, the district needs to continue working through the possible causes of weak correlation, conducting additional research as necessary. As discussed, the teacher has the greatest impact on student achievement. In order to inform decisions, there needs to be an understanding of the characteristics of quality teaching. In developing her specific and detailed outline for evaluation in education, Danielson (1996) discussed the expressed need for teachers and stakeholders to develop a shared understanding of good practice. Danielson focused the third domain of the Framework of Teaching on instruction and the characteristics that make a teacher an effective instructor, as having an effective teacher directly creates positive growth for students. Based on those research findings, standard four should be a predictable indicator of a teacher's EVAAS scores. Since this study has not supported that theory, literature was reviewed to ensure the instrument was valid, and there was a connection between the content of the standard and teacher effectiveness. As a result, the researcher suggests inter-rater reliability needs improvement by reevaluating and recalibrating scores on the teacher evaluation rubric to ensure the standards are being rated accurately to identify effective, high-quality teachers.

Research Question 6. What is the relationship between a teacher's score for reflection on his/her practice as defined by the state teacher evaluation instrument and the rating d

etermined from value-added data?

Analysis of Research Question 6, the relationship between standard five, a

teacher's ability to reflect on his/her practice as defined by the state teacher evaluation instrument, and the rating determined from value-added data, provides evidence of a statistically significant negligible, positive correlation.

Standard five evaluates teachers' abilities to reflect on their practices. As summarized in previous standards, teachers are responsible for continuously gathering data about student learning through a variety of assessments; however, it is an expectation of quality teachers to continue analyzing that data in order to make appropriate changes to instruction as determined by the needs of the students. Data are reviewed to understand learning to make decisions that ultimately impact and positively improve student performance. Effective teachers participate in professional development opportunities that support their professional and personal development as an educator (DPI, 2012c).

Standard five has a significant impact on the previous four standards. If a teacher is reflecting on his/her practice and analyzing data effectively, his/her performance should continuously improve in the previously discussed areas. As a result, if assessed appropriately, standard five should have a measurable relationship with EVAAS data. However, as the correlation was found to be negligible, the researcher reviewed the literature to ensure research validated the teacher evaluation instrument and supported the inclusion of reflection as a characteristic of quality teaching. As a result, the researcher recommends the district focus improvements on inter-rater reliability through evaluator training. The district indicators and elements should be reviewed with administrators to ensure evaluators are capturing quality teachers in their evaluations.

Recommendations

It is important for the district to continue standardizing the teacher evaluation

rubric. The district has already developed key indicators of performance to aid administrators in assigning appropriate ratings to teachers. As the district continues to strive to improve their evaluation practices, it would be beneficial to continue providing professional development opportunities for administrators. Additionally, leader supervisors need to provide feedback around data collected from the measures. Evaluators could compare specific teacher rubrics, overall mean scores, and EVAAS correlations to continue to develop their practice. Experienced evaluators should also be available to help novice administrators walk through the process and make themselves available beyond initial training opportunities. The district should also develop strategies around historical challenges or common scenarios that evaluators face. As a result of the findings in *The Widget Effect* report, TNTP (2009) recommended a reform of current evaluation practices including suggestions to ensure evaluations systems were fair, accurate, and credible in their ability to identify effective teachers and their ability to positively impact student achievement. To achieve that goal, *The Widget Effect* report recommended rigorous training and accountability for administrators (TNTP, 2009).

This study only adds to the collection of literature that discusses the importance of rigorous training and professional development of evaluators. In the RAND study of evaluation practices among 32 school districts in the United States, researchers suggested that based on the findings of the study, one key challenge in the evaluation process is the lack of training for evaluators. The study found that the evaluation models being used were standardized and focused on development and reflection, like the evaluation rubric in this study. As they defined the challenges, the RAND group provided 12 recommendations which included providing more appropriate resources and training to empower administrators to evaluate accurately and effectively. Additionally, the

recommendations also suggested that districts be more proactive in monitoring the quality of evaluators through the inclusion of expert teachers in the evaluation process (Wise et al., 1984). The variance in correlations based on school level, gender, ethnicity, and career status provides even more evidence to support the problem that results from a lack of inter-rater reliability. Principals are responsible for the development of their employees but receive minimal training around the coaching process, as the state only recommends between approximately 7 and 13 hours of training (DPI, 2012d). School administrators should be equipped to be instructional leaders at their site, including being adequately trained on recognizing effective teaching and conducting teacher evaluations.

Heneman et al. (2006) identified a positive relationship between teacher evaluation scores and student achievement gains with stronger correlations in two of the four studies. The researchers reported the cause of the higher correlation was the use of multiple, highly trained evaluators and a shared knowledge of the characteristics of good teaching. There was only a single evaluator at the locations with the lower correlations; and at those sites, the evaluators participated in less training. The report indicated that there was potential for a meaningful connection between what the teachers were observed to be doing in their classrooms and the achievement gains of their students when evaluators were rigorously trained.

Similarly, the findings from the MET study caused researchers to make recommendations for training and assessment of the evaluation processes. The report outlined suggestions for school districts to adhere more specifically to the rubric guidelines when conducting classroom observations to ensure validity, reliability, and accuracy. To do that, the report recommended that evaluators participate in training and assessment on differentiating performance. As part of that recommendation, the MET

study researchers specifically noted that all trainers in Hillsborough County had gone through professional development to learn about the process and how to appropriately utilize the tool and found that, as a result, observations were more consistent and ratings were similar when compared with a variety of observers (Bill and Melinda Gates Foundation, 2013). Additionally, they should be provided continuous opportunities for practice after initial training sessions.

In Hillsborough County, evaluators participate in a four-step training process that they consider to be a comprehensive model that prepares individuals to evaluate accurately and effectively. The four-step process includes a 6-hour online training, 3 days of in-person training, 1 day of shared observations, and conducting observations in school settings one-on-one with a trainer. In addition, once they have participated in the training program, evaluators are required to regularly calibrate their observations to improve inter-rater reliability. Hillsborough County partners with Educational Impact, an online professional development provider, to conduct their online courses for evaluators (Hillsborough County Public Schools, 2011).

Implications for Future Research

Based on the findings from this study, there are many implications for future research. Within the same parameters, a study should be conducted to analyze the individual evaluators more specifically. If patterns among evaluators could be isolated, a follow-up study has the potential to provide additional information to the district about evaluators who have a stronger correlation between observations and value-added data. If research identified evaluators within the system who were more accurately and efficiently identifying the effective teachers, results could have a significant impact on the implementation of the tools.

Multiple measures. The state in which this study was undertaken is working towards creating an overall effectiveness rating for teachers based on their ratings on standards one through five and performance as measured by standard six. The state proposes to provide each teacher with a rating of Not Effective, Effective, or Highly Effective based on their ratings. However, based on the findings from this study, the two measures are not appropriately identifying the same teachers as effective. In order to implement an overall score, the state would need to continue developing the use of the educator effectiveness rubric and to consider the use of additional multiple measures of teacher evaluation. As discussed, several studies have discussed the importance of utilizing multiple measures of teacher evaluation. As there is not a strong relationship between these two measures, the researcher supports the continued discussion of additional tools to truly capture the effectiveness of a teacher.

The MET study conducted through the Bill and Melinda Gates Foundation (2013) stressed that weights of measures of teacher effectiveness should be balanced to prevent focus on any single measure, supporting the idea of utilizing multiple measures for evaluative purposes. In the study, researchers recognized that the measures that were weighted the highest were valued the most. It was suggested that between 33% and 50% of a teacher's overall rating be determined by student growth and achievement measures. Maintaining the balance ensures that there is not a narrow focus on any single measure but encourages improvement on all measures (Bill and Melinda Gates Foundation, 2013).

The findings from the MET study also support the use of multiple observers; for example, the use of content area experts or peer observers. Reliability increased when observations of a single teacher were conducted by multiple evaluators on multiple lessons of a single teacher. When two lessons were observed by the same observer

instead of a single observation, reliability increased from .51 to .58 (based on a 0 to 1 scale). When two lessons were observed by two different evaluators, reliability increased to .67 (Bill and Melinda Gates Foundation, 2013). In addition to the principal observation and value-added data utilized for this study, research supports the use of content area expert observations to increase the reliability of the evaluation.

Additional correlations. The additional correlations calculated provided some implications for future research around the relationship between the teacher evaluation rubric scores and EVAAS data.

When disaggregated into school level, Spearman's rank-order coefficient identified positive relationships for every standard. Among elementary school teachers, the strongest relationship was found with standard four; however, for middle and high school teachers, the strongest relationship was found with standard two. One potential cause of a stronger correlation with standard four among elementary school teachers is the focus on development and pedagogy. In elementary schools, students are taught more process- and skills-based learning than significant amounts of content. Future research could further delve into the specific subjects taught by elementary school teachers to determine whether one subject has a higher correlation than another.

Findings based on career status also yielded varying results. Although both groups of teachers had the highest correlation between EVAAS and standard two, the groups had lower correlations in two different standards. For probationary teachers, a lower correlation was found with standard three, supporting the idea that novice teachers need time to develop their understanding of the content. Career teachers had the lowest correlation with standard one, teachers demonstrate leadership. One would assume that more experienced teachers would be more equipped to take on leadership responsibilities.

The research shows, however, that ratings provided for leadership roles had a weaker relationship than the other four standards. Overall, there was only a slight variation in the correlation coefficients for career and probationary teachers from the findings that included teachers from the entire sample.

In the study, the strongest relationship was found between EVAAS and standard four among Asian teachers, where $r_s=0.47$. A positive, moderate relationship was also found among Hispanic teachers for standard four, as $r_s=0.31$. Upon further research, there is limited research that investigates the effect of teacher ethnicity on principal observation ratings, as studies primarily focus on student ethnicity. Initially, a deeper delve into that specific correlation could yield more findings around the ability to effectively predict effectiveness as determined by EVAAS scores by accurately measuring a teacher's ability to facilitate learning for their students. It would be beneficial to study the subjects taught by those specific teachers and how their ratings compared to district means. Additionally, a future study that isolates the evaluators who assessed those particular employees to find if a pattern exists among specific administrators could prove to be a significant data point for decision making in the district.

Additionally, future research should also be conducted to investigate the background of evaluators. The researcher hypothesizes that correlations would be stronger among evaluators who have experience teaching the subjects they are evaluating; for example, elementary school principals evaluating elementary school teachers or principals with high school math experience evaluating high school math teachers. Based on the findings from the additional correlations, the researcher would also suggest that correlations be found for each individual school and evaluator. The

more specific information would provide the district additional data to further understand how effectively schools and administrators are identifying effective teachers. The information could be used to create professional development sessions and guide inter-rater reliability training.

As with any initiative or change, evaluation is a critical part of the process. As the district utilizes data to make changes to its evaluation processes or develop training to support the effective use of the rubric, correlations should be recalculated to measure the success of the change. The findings from this study could be used as a benchmark for the district to make continuous improvement. As rigorous training is changed and implemented, it would be necessary to recalculate the relationships between EVAAS and standards one through five the following school year.

Limitations

In conducting this study, one limitation that impacts the results lies in the inter-rater reliability, or the degree of agreement among raters. The educator evaluation system was designed with elements and descriptors to guide evaluators in selecting appropriate ratings for teachers. However, because of the human selection process, the rubric remains subjective. Because there are multiple evaluators at each of the 160 schools in the district, there is an inevitable variance in rubric scores. All evaluators participate in training and orientation sessions provided by the district; however, bias still exists as human observers will not necessarily interpret results the same way.

Summary

The relationship between teacher evaluation based on principal observation and EVAAS data in a large, urban district in a southeastern state was assessed in this study. Based on the findings, further research needs to be conducted to better understand the

causality of negligible and weak correlations to better inform the process. This study provides implications about current practice but does little to aid evaluators in using the teacher evaluation rubric to identify the teachers with the strongest value-added scores based on student achievement data. As the teacher still has the greatest impact on student achievement (McCaffrey et al., 2003), stakeholders should continue to strive to improve the processes in which highly effective teachers are identified. An effective evaluation process guides the recruitment, retention, and rewarding of quality teachers within the district.

The results of Spearman's rank-order coefficient indicated that there was not a strong correlation between the teacher evaluation instrument and EVAAS scores. However, as the quality of a teacher has the greatest impact on student achievement, focus must remain on improving teacher performance. The study aimed to compare methods used to assess quality teaching; and after analyzing the evaluation data of teachers in a large, urban school district in a southeastern state, the challenge continues to be how to identify measures that can be used to accurately evaluate educator effectiveness (Partee, 2012). There is currently not a single instrument that stakeholders agree would quickly and accurately assesses teacher effectiveness (American Federation of Teachers, 2011); however, as theories continue developing about effective teaching, there is a growing necessity for evaluation systems and processes to include multiple measures that assess the effect teachers have on student growth, performance, and success.

References

- ACT. (2011). *ACT profile report: Graduating class 2011, national*. Retrieved from www.act.org/newsroom/data/2011/pdf/profile/National2011.pdf
- Alicias, E. (2005). Toward an objective evaluation of teacher performance: The use of variance partitioning analysis. *Education Policy Analysis Archives*, 13(30), 1-13.
- American Federation of Teachers. (2011). *A guide for developing multiple measures for teacher development and evaluation*. Retrieved from http://api.ning.com/files/eC7i*SCn8hFqZT7*v*C9ib3*YbWDMkKX-D6dCHsAzwx6RZ66qy5mbt6maxQ2R-L1p4r36N7hA9sRpcsG8Y7sj0Pl50qw5jX/AFT_GUIDE.pdf
- American Recovery and Reinvestment Act of 2009. (2009). Public Law 111-5, Section 14005-6, Title XIV.
- Anderman, E., Anderman, L., Yough, M., & Gimbert, B. (2010). Value-added models of assessment: Implications for motivation and accountability. *Educational Psychologist*, 45(2), 123-137.
- Bill and Melinda Gates Foundation (2013). Measures of effective teaching project releases final research report: Findings help inform design and implementation of high-quality feedback and evaluation systems. Retrieved from <http://www.gatesfoundation.org/media-center/press-releases/2013/01/measures-of-effective-teaching-project-releases-final-research-report>
- Blumberg, A. (1985). Where we came from: Notes on supervision in the 1840's. *Journal of Curriculum and Supervision*, 1(1), 56-65.
- Briggs, D., & Domingue, B. (2011). *Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles unified school district teachers by the "Los Angeles times."* National Education Policy Center. Retrieved from <http://files.eric.ed.gov/fulltext/ED516008.pdf>
- Bruce, R., & Hoehn, L. (1980, December). *Supervisory practice in Georgia and Ohio*. Paper presented at the Annual Meeting of the Council of Professors of Instructional Supervision, Hollywood, FL.
- Burke, P., & Krey, R. (2005). *Supervision: A guide to instructional leadership* (2nd ed.). Springfield, IL: Thomas.
- Carnevale, A., Smith, N., & Strohl, J. (2010). *Help wanted: Projections of jobs and economic requirements through 2018*. Washington, DC: Georgetown Center on Education and the Workforce.

- Cogan, M. (1973). *Clinical supervision*. Boston: Houghton Mifflin.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Coleman, E. (1945). The “supervisory visit.” *Educational Leadership*, 2(4), 164-167.
- The Conference Board, Corporate Voices for Working Families, The Partnership for 21st Century Skills, and the Society for Human Resource Management. (2006, September 6). *Are they really ready to work?* Retrieved from www.p21.org/storage/documents/FINAL_REPORT_PDF09-29-06.pdf
- Corcoran, S. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Executive Summary. Education Policy for Action Series. Annenberg Institute for School Reform at Brown University.
- Creswell, J. (2008). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- Cubberley, E. (1929). *Public school administration* (3rd ed.). Boston: Houghton Mifflin.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Department of Public Instruction. (2012a). *McREL: North Carolina educator evaluation system*. Retrieved from mxweb3.media-x.com/home/ncval/
- Department of Public Instruction. (2012b). *Mission statement*. Retrieved from www.ncpublicschools.org/organization/mission/
- Department of Public Instruction. (2012c). *North Carolina resources*. Retrieved from www.ncpublicschools.org/profdev/training/online-evaluation/
- Department of Public Instruction. (2012d). *North Carolina teacher evaluation*. Retrieved from www.ncpublicschools.org/profdev/training/teacher/
- Department of Public Instruction. (2012e). North Carolina teacher evaluation process: Teacher evaluation manual. Retrieved from www.ncpublicschools.org/docs/effectiveness-model/ncees/instruments/teach-eval-manual.pdf

- Department of Public Instruction. (2013a). EVAAS. Retrieved from www.ncpublicschools.org/effectiveness-model/evaas/?&print=true
- Department of Public Instruction. (2013b). *EVAAS selection*. Retrieved from www.ncpublicschools.org/effectiveness-model/evaas/selection/?&print=true
- Department of Public Instruction. (2013c). *Measuring growth for educator effectiveness: A guide to the use of student growth data in the evaluation of North Carolina teachers*. Retrieved from www.ncpublicschools.org/docs/effectiveness-model/ncees/measure-growth-guide.pdf
- Department of Public Instruction. (2013d). *North Carolina educator evaluation model*. Retrieved from www.ncpublicschools.org/effectiveness-model/ncees/
- Department of Public Instruction. (2013e). *North Carolina's race to the top*. Retrieved from www.ncpublicschools.org/rttt/
- Department of Public Instruction (2013f). *Facts and figures 2012-2013*. Retrieved from <http://www.ncpublicschools.org/docs/fbs/resources/data/factsfigures/2012-13figures.pdf>
- Dewey, J. (1981). *The philosophy of John Dewey* (J. McDermott, Ed.). Chicago, IL: University of Chicago Press.
- Duncan, A., & Martin, C. (2010). *ESEA blueprint for reform*. Washington, DC: US Department of Education, Office of Planning, Evaluation and Policy Development.
- Fehr, S. (2001, August). *The role of educational supervision in the United States public schools from 1970 to 2000 as reflected in the supervision literature*. Unpublished doctoral dissertation, Pennsylvania State University, State College.
- Glatthorn, A. (1984). *Differentiated supervision*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Glickman, C., Gordon, S., & Ross-Gordon, J. (1998). *Supervision of instruction: A developmental approach* (4th ed.). Boston: Allyn & Bacon.
- Goldhammer, R., Anderson, R., & Krajewski, R. (1993). *Clinical supervision: Special methods for the supervision of teachers*. Orlando, FL: Harcourt Brace College Publishers.
- Guarino, C., Reckase, M., & Wooldridge, J. (2011). *Evaluating value-added methods of estimating of teacher performance*. Society for Research on Educational Effectiveness. Retrieved from <http://files.eric.ed.gov/fulltext/ED518225.pdf>

- Heneman, H., Milanowski, A., Kimball, S., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge-and skill-based pay*. Philadelphia, PA: University of Pennsylvania, Graduate School of Education, Consortium for Policy Research in Education. DOI: 10.12698/cpre.2013.rb45
- Herman, J., Heritage, M., & Goldschmidt, P. (2011). *Developing and selecting assessments of student growth for use in teacher evaluation systems*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from <https://www.cse.ucla.edu/products/policy/TestScoresTeacherEval.pdf>
- Hillsborough County Public Schools. (2011). *Teacher evaluation handbook: Empowering effective teachers*. Retrieved from http://www.education-first.com/files/Hillsborough_Teacher_Evaluation_Handbook.pdf
- Hunter, M. (1984). Knowing, teaching, and supervising. *Using what we know about teaching* (pp. 169-192). Alexandria, VA: Association for Supervision and Curriculum Development.
- Isenberg, E., & Hock, H. (2011). *Design of value-added models for impact and team in DC public schools, 2010-2011 school year, final report*. Mathematica Policy Research, Inc. Retrieved from <http://dcps.dc.gov/DCPS/Files/downloads/In-the-Classroom/Design%20of%20Value-Added%20Models%20for%20DCPS%202010-2011.pdf>
- Lewis, H., & Leps, J. (1946). When principals supervise. *Educational Leadership*, 3(4), 160-163.
- Lipscomb, S., Chiang, H., & Gill, B. (2012). *Value-added estimates for phase 1 of the Pennsylvania teacher and principal evaluation pilot*. Executive Summary. Mathematica Policy Research, Inc. Retrieved from http://www.mathematica-mpr.com/~media/publications/PDFs/education/value-added_PAteachprin_tchrpt.pdf
- Lloyd, J. (2009). *Instituting a value added assessment system in Ohio: The professional development implications from the perspective of the district value-added specialists*. Dissertation Abstracts International Section A, 69. Retrieved from https://etd.ohiolink.edu/rws_etd/document/get/ashland1226675998/inline
- Lund Research. (2013). *Spearman's correlation in SPSS*. Retrieved from <https://statistics.laerd.com/premium/sroc/spearmans-rank-order-correlation-in-spss.php>
- Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: ASCD.

- Marzano, R., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McCaffrey, J., Lockwood, D., Koretz, D., & Hamilton, L. (2003). *Evaluating value added models for teacher accountability*. Santa Monica, CA: RAND Corporation. Retrieved from www.rand.org/pubs/monographs/2004/RAND_MG158.pdf
- McGreal, T. (1983). *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Melchoir, W. (1950). *Instructional supervision: A guide to modern practice*. Boston: Heath.
- Millard, J. (1976). *Teacher evaluation*. Ankeny, IA: Heartland Educational Agency.
- National Board for Professional Teaching Standards. (2013). *National board standards*. Retrieved from www.nbpts.org/national-board-standards
- National Center for Education Statistics. (2012, June 13). Digest of Education Statistics, 2011. Retrieved from nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012001
- National Council on Teacher Quality. (2013, October). State of the states 2013 connecting the dots: Using evaluations of teacher effectiveness to inform policy and practice. Retrieved from www.nctq.org/dmsStage/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report
- The New Teacher Project. (2009). *The Widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.
- Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23), 1-22.
- Nicholson, M., & Brown, J. (2010). Value-added analysis in instruction. *School Administrator*, 67(2), 26-30.
- No Child Left Behind Act. (2001). Public Law No. 107-110, 115 Stat. 1425. (2002).
- O'Malley, K., McClarty, K., Magda, T., & Burling, K. (2011, April). *Making sense of the metrics: Student growth, value-added models, and teacher effectiveness*. Pearson Assessments, 19. Retrieved from <http://researchnetwork.pearson.com/wp-content/uploads/F143689371.pdf>

- The Organisation for Economic Co-operation and Development. (2010). *Education at a glance 2010: OECD indicators*. Retrieved from www.oecd.org/education/skills-beyond-school/educationataglance2010oecdindicators.htm
- Partee, G. (2012). *Using multiple evaluation measures to improve teacher effectiveness: State strategies from round 2 of no child left behind act waivers*. Retrieved from www.americanprogress.org/issues/education/report/2012/12/18/48368/using-multiple-evaluation-measures-to-improve-teacher-effectiveness/
- Partnership for 21st Century Skills. (2011, March). *Framework for 21st century learning*. Retrieved from <http://www.p21.org/our-work/p21-framework>
- Peterson, K., & Kauchak, D. (1982). *Teacher evaluation: Perspectives, practices, and promises*. Retrieved from <http://files.eric.ed.gov/fulltext/ED233996.pdf>
- Reich, R. (1996, May 22). *Keynote address to the conference on the economic war among the states national academy of sciences*. United States Department of Labor. Retrieved from www.dol.gov/dol/aboutdol/history/reich/speeches/sp960522.htm
- Sanders, W., & Horn, S. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299-311. Retrieved from [http://addingvalue.wceruw.org/RelatedBibliography/Articles/Sanders & Horn.pdf](http://addingvalue.wceruw.org/RelatedBibliography/Articles/Sanders%20&%20Horn.pdf)
- Sartain, L., Stoelinga, S., & Brown, E. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Research Report. Consortium on Chicago School Research. Retrieved from <https://ccsr.uchicago.edu/sites/default/files/publications/Teacher%20Eval%20Report%20FINAL.pdf>
- School District. (2010, February 8). *Other news*. Retrieved from www.cms.k12.nc.us/cmsdepartments/accountability/TEP/Pages/Othernews.aspx
- School District. (2010, May 5). *Pay for performance begins*. Retrieved from www.cms.k12.nc.us/stories/Pages/Pay-for-performanceworkbegins.aspx
- School District. (2012, March 6). *Talent effectiveness project*. Retrieved from www.cms.k12.nc.us/cmsdepartments/accountability/TEP/Pages/TEP%20Overview.aspx
- School District. (2013a). *Did you know?* Retrieved from www.cms.k12.nc.us/mediaroom/aboutus/Pages/Didyouknow.aspx
- School District. (2013b). *Fast facts*. Retrieved from www.cms.k12.nc.us/mediaroom/aboutus/Pages/FastFacts.aspx

- School District. (2013, January 18). *Focus groups will address compensation design*. Retrieved from www.cms.k12.nc.us/News/Pages/Focusgroupswilladdresscompensationdesign.aspx
- School District Project. (2013). *About project LIFT*. Retrieved from www.projectliftcharlotte.org
- Tracy, S. (1995, May/June). How historical concepts of supervision relate to supervisory practices today. *The Clearing House*, 68(5), 320-324.
- Tucker, P., & Stronge, J. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Value-added teacher evaluation. (2011). *American School Board Journal*, 198(4), 57-58.
- Vanderbilt University. (2008). *Value-added and other methods for measuring school performance: An analysis of performance measurement strategies in teacher incentive fund proposals*. Research Brief. National Center on Performance Incentives. Retrieved from https://my.vanderbilt.edu/performanceincentives/files/2012/10/200817_MeyerChristian_ValueAdded.pdf
- Weerasinghe, D., Orsak, T., & Mendro, R. (1997). *Value added productivity indicators: A statistical comparison of the pre-test/post-test model and gain model*. <http://files.eric.ed.gov/fulltext/ED411245.pdf>
- WestEd. (2012). *Validating North Carolina's educator evaluation system: Evidence in chain of claims to support system use*. Retrieved from <http://www.ncpublicschools.org/docs/effectiveness-model/evaas/selection/validating-evaluation.pdf>
- Wetzel, W. (1929, February). Scientific supervision and curriculum building. *The School Review*, 37(2), 179-192.
- Whitehead, M. (1952). Teachers look at supervision. *Educational Leadership*, 10(2), 1011-1106.
- Wise, E., Darling-Hammond, L., McLaughlin, M., & Bernstein, H. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: RAND.

Appendix A

State Teacher Evaluation Rubric

Rubric for Evaluating North Carolina Teachers (Required)

This form should be used for the teacher self-assessment, classroom observation, and the summary evaluation.

Name: _____ Date: _____

School: _____ District: _____

Evaluator: _____ Title: _____

Start Time: _____ End Time: _____

Standard I: Teachers demonstrate leadership

a. Teachers lead in their classrooms. Teachers demonstrate leadership by taking responsibility for the progress of all students to ensure that they graduate from high school, are globally competitive for work and postsecondary education, and are prepared for life in the 21st century. Teachers communicate this vision to their students. Using a variety of data sources, they organize, plan, and set goals that meet the needs of the individual student and the class. Teachers use various types of assessment data during the school year to evaluate student progress and to make adjustments to the teaching and learning process. They establish a safe, orderly environment, and create a culture that empowers students to collaborate and become lifelong learners.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Understands how they contribute to students graduating from high school. <input type="checkbox"/> Uses data to understand the skills and abilities of students.	... and <input type="checkbox"/> Takes responsibility for the progress of students to ensure that they graduate from high school. <input type="checkbox"/> Provides evidence of data driven instruction throughout all classroom activities. <input type="checkbox"/> Establishes a safe and orderly classroom.	... and <input type="checkbox"/> Communicates to students the vision of being prepared for life in the 21st century. <input type="checkbox"/> Evaluates student progress using a variety of assessment data. <input type="checkbox"/> Creates a classroom culture that empowers students to collaborate.	... and <input type="checkbox"/> Encourages students to take responsibility for their own learning. <input type="checkbox"/> Uses classroom assessment data to inform program planning. <input type="checkbox"/> Empowers and encourages students to create and maintain a safe and supportive school and community environment.	
b. Teachers demonstrate leadership in the school. Teachers work collaboratively with school personnel to create a professional learning community. They analyze and use local, state, and national data to develop goals and strategies in the school improvement plan that enhances student learning and teacher working conditions. Teachers provide input in determining the school budget and in the selection of professional development that meets the needs of students and their own professional growth. They participate in the hiring process and collaborate with their colleagues to mentor and support teachers to improve the effectiveness of their departments or grade levels.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Attends professional learning community meetings. <input type="checkbox"/> Displays	... and <input type="checkbox"/> Participates in professional learning community. <input type="checkbox"/> Participates in	... and <input type="checkbox"/> Assumes a leadership role in professional learning community.	... and <input type="checkbox"/> Collaborates with colleagues to improve the quality of learning in the school.	

awareness of the goals of the school improvement plan.	developing and/or implementing the school improvement plan.	<input type="checkbox"/> Collaborates with school personnel on school improvement activities.	<input type="checkbox"/> Assumes a leadership role in implementing school improvement plan throughout the building.	
c. Teachers lead the teaching profession. Teachers strive to improve the teaching profession. They contribute to the establishment of positive working conditions in their school. They actively participate in and advocate for decision-making structures in education and government that take advantage of the expertise of teachers. Teachers promote professional growth for all educators and collaborate with their colleagues to improve the profession.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Has knowledge of opportunities and the need for professional growth and begins to establish relationships with colleagues.	... and Contributes to the: <input type="checkbox"/> Improvement of the profession through professional growth. <input type="checkbox"/> Establishment of positive working relationships <input type="checkbox"/> School's decision-making processes as required.	... and <input type="checkbox"/> Promotes positive working relationships through professional growth activities and collaboration.	... and <input type="checkbox"/> Seeks opportunities to lead professional growth activities and decision-making processes.	
d. Teachers advocate for schools and students. Teachers advocate for positive change in policies and practices affecting student learning. They participate in the implementation of initiatives to improve the education of students.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Knows about the policies and practices affecting student learning.	... and <input type="checkbox"/> Supports positive change in policies and practices affecting student learning.	... and <input type="checkbox"/> Participates in developing policies and practices to improve student learning.	... and <input type="checkbox"/> Actively participates, promotes, and provides strong supporting evidence for implementation of initiatives to improve education.	
e. Teachers demonstrate high ethical standards. Teachers demonstrate ethical principles including honesty, integrity, fair treatment, and respect for others. Teachers uphold the Code of Ethics for North Carolina Educators (effective June 1, 1997) and the <i>Standards for Professional Conduct</i> adopted April 1, 1998. (www.ncptsc.org)				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Understands the importance of ethical behavior as outlined in the <i>Code of Ethics for North Carolina Educators</i> and the <i>Standards for Professional</i>	... and <input type="checkbox"/> Demonstrates ethical behavior through adherence to the <i>Code of Ethics for North Carolina</i>	... and <input type="checkbox"/> Knows and upholds the <i>Code of Ethics for North Carolina Educators</i> and the <i>Standards</i>	... and <input type="checkbox"/> Models the tenets of the <i>Code of Ethics for North Carolina Educators</i> and the <i>Standards</i>	

<i>Conduct.</i>	<i>Educators and the Standards for Professional Conduct.</i>	<i>for Professional Conduct.</i>	<i>for Professional Conduct and encourages others to do the same.</i>	
Comments				

Examples of Artifacts:

- Lessonplans
- Journals
- Student handbooks
- Student work
- School improvement planning
- Service on committees
- Relevant data
- Class rules and procedures
- Participation in The Teacher Working Condition Survey
- Professional Learning Communities
- Membership in professional organizations
- Formal and informal mentoring
- Surveys
- National Board Certification
- Discipline record

Standard II: Teachers Establish a Respectful Environment for a Diverse Population of Students

a. Teachers provide an environment in which each child has a positive, nurturing relationship with caring adults. Teachers encourage an environment that is inviting, respectful, supportive, inclusive, and flexible.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Appreciates and understands the need to establish nurturing relationships.	... and <input type="checkbox"/> Establishes an inviting, respectful, inclusive, flexible, and supportive learning environment.	... and <input type="checkbox"/> Maintains a positive and nurturing learning environment.	... and <input type="checkbox"/> Encourages and advises others to provide a nurturing and positive learning environment for all students.	
b. Teachers embrace diversity in the school community and in the world. Teachers demonstrate their knowledge of the history of diverse cultures and their role in shaping global issues. They actively select materials and develop lessons that counteract stereotypes and incorporate histories and contributions of all cultures. Teachers recognize the influence of race, ethnicity, gender, religion, and other aspects of culture on a student's development and personality. Teachers strive to understand how a student's culture and background may influence his or her school performance. Teachers consider and incorporate different points of view in their instruction.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Acknowledges that diverse cultures impact the world. <input type="checkbox"/> Demonstrates awareness of the diversity of students in the	... and <input type="checkbox"/> Displays knowledge of diverse cultures, their histories, and their roles in shaping global issues.	... and <input type="checkbox"/> Uses materials or lessons that counteract stereotypes and acknowledged	... and <input type="checkbox"/> Promotes a deep understanding of cultures through the integration of culturally sensitive	

classroom.	<input type="checkbox"/> Acknowledges the influence of race, ethnicity, gender, religion, socio-economics, and culture on a student's development and attitudes.	<p>ges the contributions of all cultures.</p> <input type="checkbox"/> Consistently incorporates different points of view in instruction.	<p>materials and ideas throughout the curriculum.</p> <input type="checkbox"/> Capitalizes on diversity as an asset in the classroom.	
c. Teachers treat students as individuals. Teachers maintain high expectations, including graduation from high school, for students of all backgrounds. Teachers appreciate the differences and value the contributions of each student in the learning environment by building positive, appropriate relationships.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Holds high expectations of students.	<p>... and</p> <input type="checkbox"/> Communicates high expectations for all students.	<p>... and</p> <input type="checkbox"/> Encourages and values contributions of students, regardless of background or ability.	<p>... and</p> <input type="checkbox"/> Helps students hold high expectations for themselves and their peers.	
d. Teachers adapt their teaching for the benefit of students with special needs. Teachers collaborate with the range of support specialists to help meet the special needs of all students. Through inclusion and other models of effective practice, teachers engage students to ensure that their needs are met.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Recognizes that students have a variety of learning needs. <input type="checkbox"/> Is knowledgeable of effective practices for students with special needs.	<p>... and</p> <input type="checkbox"/> Collaborates with specialists who can support the special learning needs of students. <input type="checkbox"/> Provides unique learning opportunities such as inclusion and research based effective practices for students with special needs.	<p>... and</p> <input type="checkbox"/> Understands the roles of and collaborates with the full range of support specialists to help meet the special needs of all students. <input type="checkbox"/> Effectively engages special needs students in learning activities and ensures their unique learning needs are met.	<p>... and</p> <input type="checkbox"/> Anticipates the unique learning needs of students and solicits assistance from within and outside the school to address those needs. <input type="checkbox"/> Adapts instruction for the benefit of students with special needs and helps colleagues do the same for their students.	
e. Teachers work collaboratively with the families and significant adults in the lives of their students. Teachers recognize that educating children is a shared responsibility involving the school, parents or guardians, and the community. Teachers improve communication and collaboration between the school and the home and community in order to promote trust and understanding and build partnerships with all segments of the school community. Teachers seek solutions to overcome cultural and economic obstacles that may stand in the way of effective family and community involvement in the education of their students.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Responds to family and community concerns.	<p>... and</p> <input type="checkbox"/> Communicates and collaborates with the	<p>... and</p> <input type="checkbox"/> Recognizes obstacles to family	<p>... and</p> <input type="checkbox"/> Promotes trust and	

	home and community for the benefit of students.	and community participation and conscientiously seeks solutions to overcome them.	understanding throughout the school community.	
Comments				

Examples of Artifacts

- Student profiles
- Student surveys
- Cooperation with ESL teachers
- Lessons that integrate international content
- Documentation of referral data and use of IEPs
- Communications with parents/ community
- Professional development on cultural attitudes and awareness
- Use of technology to incorporate cultural awareness into lessons

Standard III: Teachers Know the Content They Teach

<p>a. Teachers align their instruction with the <i>North Carolina Standard Course of Study</i>. In order to enhance the <i>North Carolina Standard Course of Study</i>, teachers investigate the content standards developed by professional organizations in their specialty area. They develop and apply strategies to make the curriculum rigorous and relevant for all students and provide a balanced curriculum that enhances literacy skills. Elementary teachers have explicit and thorough preparation in literacy instruction. Middle and high school teachers incorporate literacy instruction within the content area or discipline.</p>				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Demonstrates an awareness of the <i>North Carolina Standard Course of Study</i> and references it in the preparation of lesson plans. <input type="checkbox"/> <i>Elementary:</i> Begins to integrate literacy instruction in selected lessons. <input type="checkbox"/> <i>Secondary:</i> Recognizes the importance of integrating literacy strategies within the content areas.	<p>... and</p> <input type="checkbox"/> Understands the <i>North Carolina Standard Course of Study</i> , uses it in preparation of lesson plans, and applies strategies to make the curriculum rigorous and relevant. <input type="checkbox"/> <i>Elementary:</i> Integrates effective literacy instruction throughout the curriculum.	<p>... and</p> <input type="checkbox"/> Develops and applies strategies based on the <i>North Carolina Standard Course of Study</i> and standards developed by professional organizations to make the curriculum balanced, rigorous and relevant. <input type="checkbox"/> <i>Elementary:</i> Evaluates and reflects upon the effectiveness of literacy instruction. <input type="checkbox"/> <i>Secondary:</i> Evaluates and	<p>... and</p> <input type="checkbox"/> Assists colleagues in applying such strategies in their classrooms. <input type="checkbox"/> <i>Elementary:</i> Makes necessary changes to instructional practice to improve student learning. <input type="checkbox"/> <i>Secondary:</i> Makes necessary changes to instructional practice to improve student learning.	

	<input type="checkbox"/> <i>Secondary:</i> Incorporates a wide variety of literacy skills within content areas to enhance learning.	reflects upon the effectiveness of literacy instruction within content areas.		
b. Teachers know the content appropriate to their teaching specialty. Teachers bring a richness and depth of understanding to their classrooms by knowing their subjects beyond the content they are expected to teach and by directing students' natural curiosity into an interest in learning. Elementary teachers have broad knowledge across disciplines. Middle school and high school teachers have depth in one or more specific content areas or disciplines.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Demonstrates a basic level of content knowledge in the teaching specialty to which assigned.	... and <input type="checkbox"/> Demonstrates an appropriate level of content knowledge in the teaching specialty to which assigned.	... and <input type="checkbox"/> Applies knowledge of subject beyond the content in assigned teaching specialty. Motivates students to investigate the content area to expand their knowledge and satisfy their natural curiosity.	... and <input type="checkbox"/> Extends knowledge of subject beyond content in their teaching specialty and sparks students' curiosity for learning beyond the required course work.	
c. Teachers recognize the interconnectedness of content areas/disciplines. Teachers know the links and vertical alignment of the grade or subject they teach and the <i>North Carolina Standard Course of Study</i> . Teachers understand how the content they teach relates to other disciplines in order to deepen understanding and connect learning for students. Teachers promote global awareness and its relevance to subjects they teach.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Understand the links between grade/subject and the <i>North Carolina Standard Course of Study</i> . <input type="checkbox"/> Displays global awareness.	... and <input type="checkbox"/> Demonstrates knowledge of links between grade/subject and the <i>North Carolina Standard Course of Study</i> . <input type="checkbox"/> Promotes global awareness and its relevance to the subjects.	... and <input type="checkbox"/> Demonstrates knowledge of the links and vertical alignment of the grade or subject area and the <i>North Carolina Standard Course of Study</i> . Relates content to other disciplines. <input type="checkbox"/> Integrates global awareness activities throughout lesson plans and classroom	... and <input type="checkbox"/> Collaborates with teachers from other grades or subject areas to establish links between disciplines and influence school-wide curriculum and teaching practice. <input type="checkbox"/> Promotes global awareness and its relevance to all faculty members, influencing curriculum and teaching practices throughout the school.	

		instructional practices.		
<p>d. Teachers make instruction relevant to students. Teachers incorporate 21st century life skills into their teaching deliberately, strategically, and broadly. These skills include leadership, ethics, accountability, adaptability, personal productivity, personal responsibility, people skills, self-direction, and social responsibility. Teachers help their students understand the relationship between the <i>North Carolina Standard Course of Study</i> and 21st century content, which includes global awareness; financial, economic, business and entrepreneurial literacy; civic literacy; and health awareness.</p>				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Identifies relationships between the North Carolina Standard Course of Study and life in the 21st century.	... and <input type="checkbox"/> Identifies relationships between the core content and 21st century content.	... and <input type="checkbox"/> Integrates core content and 21st century content throughout lesson plans and classroom instructional practices.	... and <input type="checkbox"/> Deepens students' understandings of 21st century skills and helps them make their own connections and develop new skills.	
Comments				

Examples of Artifacts:

- Display of creative student work
- Use of NC Standard Course of Study
- Lesson plans
- Content standards

Standard IV: Teachers Facilitate Learning for Their Students

<p>a. Teachers know the ways in which learning takes place, and they know the appropriate levels of intellectual, physical, social, and emotional development of their students. Teachers know how students think and learn. Teachers understand the influences that affect individual student learning (development, culture, language proficiency, etc.) and differentiate their instruction accordingly. Teachers keep abreast of evolving research about student learning. They adapt resources to address the strengths and weaknesses of their students.</p>				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Understands developmental levels of students and recognizes the need to differentiate instruction.	... and <input type="checkbox"/> Understands developmental levels of students and appropriately differentiates instruction. <input type="checkbox"/> Assesses resources needed to address strengths and	... and <input type="checkbox"/> Identifies appropriate developmental levels of students and consistently and appropriately differentiates instruction. <input type="checkbox"/> Reviews and uses	... and <input type="checkbox"/> Encourages and guides colleagues to adapt instruction to align with students' developmental levels. <input type="checkbox"/> Stays abreast of current research about student learning and	

	weakness of students.	alternative resources or adapts existing resources to take advantage of student strengths or address weaknesses.	emerging resources and encourages the school to adopt or adapt them for the benefit of all students.	
b. Teachers plan instruction appropriate for their students. Teachers collaborate with their colleagues and use a variety of data sources for short- and long-range planning based on the North Carolina Standard Course of Study. These plans reflect an understanding of how students learn. Teachers engage students in the learning process. They understand that instructional plans must be consistently monitored and modified to enhance learning. Teachers make the curriculum responsive to cultural differences and individual learning needs.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Recognizes data sources important to planning instruction.	... and <input type="checkbox"/> Uses a variety of data for short- and long- range planning of instruction. Monitors and modifies instructional plans to enhance student learning.	... and <input type="checkbox"/> Monitors student performance and responds to individual learning needs in order to engage students in learning.	... and <input type="checkbox"/> Monitors student performance and responds to cultural diversity and learning needs through the school improvement process.	
c. Teachers use a variety of instructional methods. Teachers choose the methods and techniques that are most effective in meeting the needs of their students as they strive to eliminate achievement gaps. Teachers employ a wide range of techniques including information and communication technology, learning styles, and differentiated instruction.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Demonstrates awareness of the variety of methods and materials necessary to meet the needs of all students.	... and <input type="checkbox"/> Demonstrates awareness or use of appropriate methods and materials necessary to meet the needs of all students.	... and <input type="checkbox"/> Ensures the success of all students through the selection and utilization of appropriate methods and materials.	... and <input type="checkbox"/> Stays abreast of emerging research areas and new and innovative materials and incorporates them into lesson plans and instructional strategies.	
d. Teachers integrate and utilize technology in their instruction. Teachers know when and how to use technology to maximize student learning. Teachers help students use technology to learn content, think critically, solve problems, discern reliability, use information, communicate, innovate, and collaborate.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Assesses effective types of technology to use for instruction.	... and <input type="checkbox"/> Demonstrates knowledge of how to utilize technology in instruction.	... and <input type="checkbox"/> Integrates technology with instruction to maximize student learning.	... and <input type="checkbox"/> Provides evidence of student engagement in higher level thinking skills through the integration of	

			technology.	
<p>e. Teachers help students develop critical-thinking and problem-solving skills. Teachers encourage students to ask questions, think creatively, develop and test innovative ideas, synthesize knowledge, and draw conclusions. They help students exercise and communicate sound reasoning; understand connections; make complex choices; and frame, analyze, and solve problems.</p>				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Understands the importance of developing students' critical-thinking and problem solving skills.	... and <input type="checkbox"/> Demonstrates knowledge of processes needed to support students in acquiring critical thinking skills and problem solving skills.	... and Teaches students the processes needed to: <ul style="list-style-type: none"> <input type="checkbox"/> think creatively and critically, <input type="checkbox"/> develop and test innovative ideas, <input type="checkbox"/> synthesize knowledge, <input type="checkbox"/> draw conclusions, <input type="checkbox"/> exercise and communicate sound reasoning, <input type="checkbox"/> understand connections, <input type="checkbox"/> make complex choices, and <input type="checkbox"/> frame, analyze and solve problems. 	... and <input type="checkbox"/> Encourages and assists teachers throughout the school to integrate critical thinking and problem solving skills into their instructional practices.	
<p>f. Teachers help students work in teams and develop leadership qualities. Teachers teach the importance of cooperation and collaboration. They organize learning teams in order to help students define roles, strengthen social ties, improve communication and collaborative skills, interact with people from different cultures and backgrounds, and develop leadership qualities.</p>				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Provides opportunities for cooperation, collaboration, and leadership through student learning teams.	... and <input type="checkbox"/> Organizes student learning teams for the purpose of developing cooperation, collaboration, and student leadership.	... and <input type="checkbox"/> Encourages students to create and manage learning teams.	... and <input type="checkbox"/> Fosters the development of student leadership and teamwork skills to be used beyond the classroom.	
<p>g. Teachers communicate effectively. Teachers communicate in ways that are clearly understood by their students. They are perceptive listeners and are able to communicate with students in a variety of ways even when language is a barrier. Teachers help students articulate thoughts and ideas clearly and effectively.</p>				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Demonstrates the ability to effectively	... and <input type="checkbox"/> Uses a variety of	... and <input type="checkbox"/> Creates a variety	... and <input type="checkbox"/> Anticipates	

<p>communicate with students.</p> <p><input type="checkbox"/> Provides opportunities for students to articulate thoughts and ideas</p>	<p>methods for communication with all students.</p> <p><input type="checkbox"/> Consistently encourages and supports students to articulate thoughts and ideas clearly and effectively.</p>	<p>of methods to communicate with all students.</p> <p><input type="checkbox"/> Establishes classroom practices, which encourage all students to develop effective communication skills.</p>	<p>possible student misunderstandings and proactively develops teaching techniques to mitigate concerns.</p> <p><input type="checkbox"/> Establishes school- wide and grade appropriate vehicles to encourage students throughout the school to develop effective communication skills.</p>	
<p>h. Teachers use a variety of methods to assess what each student has learned. Teachers use multiple indicators, including formative and summative assessments, to evaluate student progress and growth as they strive to eliminate achievement gaps. Teachers provide opportunities, methods, feedback, and tools for students to assess themselves and each other. Teachers use 21st century assessment systems to inform instruction and demonstrate evidence of students' 21st century knowledge, skills, performance, and dispositions.</p>				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<p><input type="checkbox"/> Uses indicators to monitor and evaluate student progress.</p> <p><input type="checkbox"/> Assesses students in the attainment of 21st century knowledge, skills, and dispositions.</p>	<p>... and</p> <p><input type="checkbox"/> Uses multiple indicators, both formative and summative, to monitor and evaluate student progress and to inform instruction.</p> <p><input type="checkbox"/> Provides evidence that students attain 21st century knowledge, skills and dispositions.</p>	<p>... and</p> <p><input type="checkbox"/> Uses the information gained from the assessment activities to improve teaching practice and student learning.</p> <p><input type="checkbox"/> Provides opportunities for students to assess themselves and others.</p>	<p>... and</p> <p><input type="checkbox"/> Teaches students and encourages them to use peer and self-assessment feedback to assess their own learning.</p> <p><input type="checkbox"/> Encourages and guides colleagues to assess 21st century skills, knowledge, and dispositions and to use the assessment information to adjust their instructional practice.</p>	
<p>Comments</p>				

Examples of Artifacts:

- Lesson plans
- Display of technology used
- Professional development

- Using of student learning teams
- Documentation of differentiated instruction
- Materials used to promote critical thinking and problem solving
- Collaborative lesson planning

Standard V: Teachers Reflect on Their Practice

a. Teachers analyze student learning. Teachers think systematically and critically about student learning in their classrooms and schools: why learning happens and what can be done to improve achievement. Teachers collect and analyze student performance data to improve school and classroom effectiveness. They adapt their practice based on research and data to best meet the needs of students.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Recognizes the need to improve student learning in the classroom.	... and <input type="checkbox"/> Provides ideas about what can be done to improve student learning in their classroom.	... and <input type="checkbox"/> Thinks systematically and critically about learning in their classroom: Why learning happens and what can be done to improve student achievement.	... and <input type="checkbox"/> Provides a detailed analysis about what can be done to improve student learning and uses such analyses to adapt instructional practices and materials within the classroom and at the school level.	
b. Teachers link professional growth to their professional goals. Teachers participate in continued, high-quality professional development that reflects a global view of educational practices; includes 21st century skills and knowledge; aligns with the State Board of Education priorities; and meets the needs of students and their own professional growth.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Understands the importance of professional development.	... and <input type="checkbox"/> Participates in professional development aligned with professional goals.	... and <input type="checkbox"/> Participates in professional development activities aligned with goals and student needs.	... and <input type="checkbox"/> Applies and implements knowledge and skills attained from professional development consistent with its intent.	
c. Teachers function effectively in a complex, dynamic environment. Understanding that change is constant, teachers actively investigate and consider new ideas that improve teaching and learning. They adapt their practice based on research and data to best meet the needs of their students.				
Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
<input type="checkbox"/> Is knowledgeable of current research-based approaches to teaching and learning.	... and <input type="checkbox"/> Considers and uses a variety of research-based approaches to improve teaching and learning.	... and <input type="checkbox"/> Actively investigates and considers alternative research-based approaches to	... and <input type="checkbox"/> Adapts professional practice based on data and evaluates impact on	

		improve teaching and learning and uses such approaches as appropriate.	student learning.	
Comments				

Examples of Artifacts:

- Lesson plans
- Formative assessments
- Student work
- Professional growth plan
- Completion of professional development
- Participation in professional learning community
- Formative and summative assessment data

Rubric for Evaluating North Carolina Teachers Signature Page

Teacher Signature

Date

Evaluator Signature

Date

Comments Attached: _____ Yes _____ No

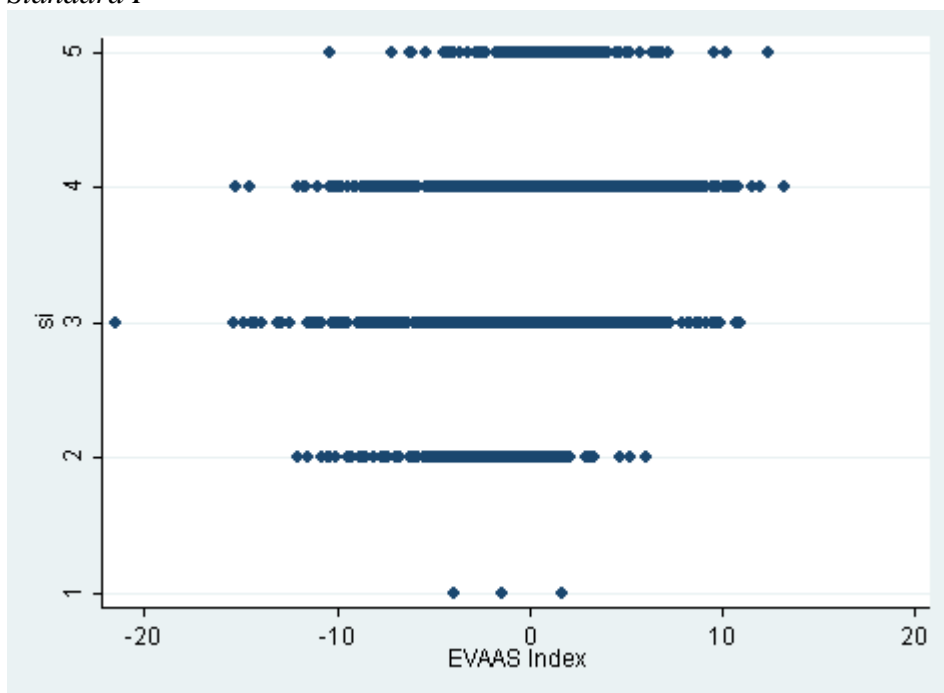
Principal/Evaluator Signature (Signature indicates question above regarding comments has been addressed)

Date

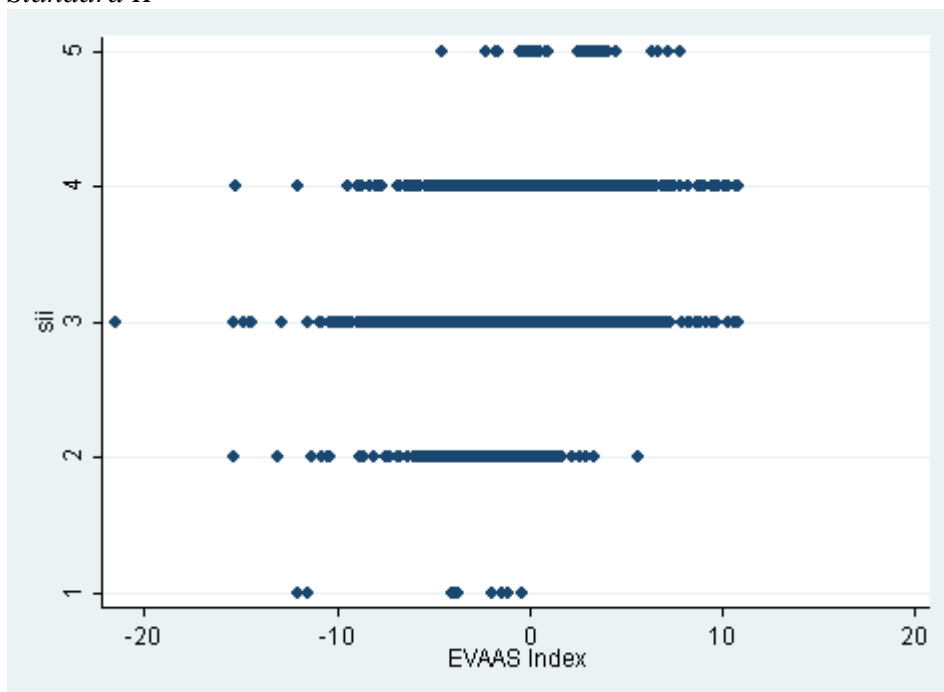
Note: The teacher's signature on this form represents neither acceptance nor approval of the report. It does, however, indicate that the teacher has reviewed the report with the evaluator and may reply in writing. The signature of the principal or evaluator verifies that the report has been reviewed and that the proper process has been followed according to North Carolina State Board of Education Policy for the Teacher Evaluation Process.

Appendix B
Data Scatterplots

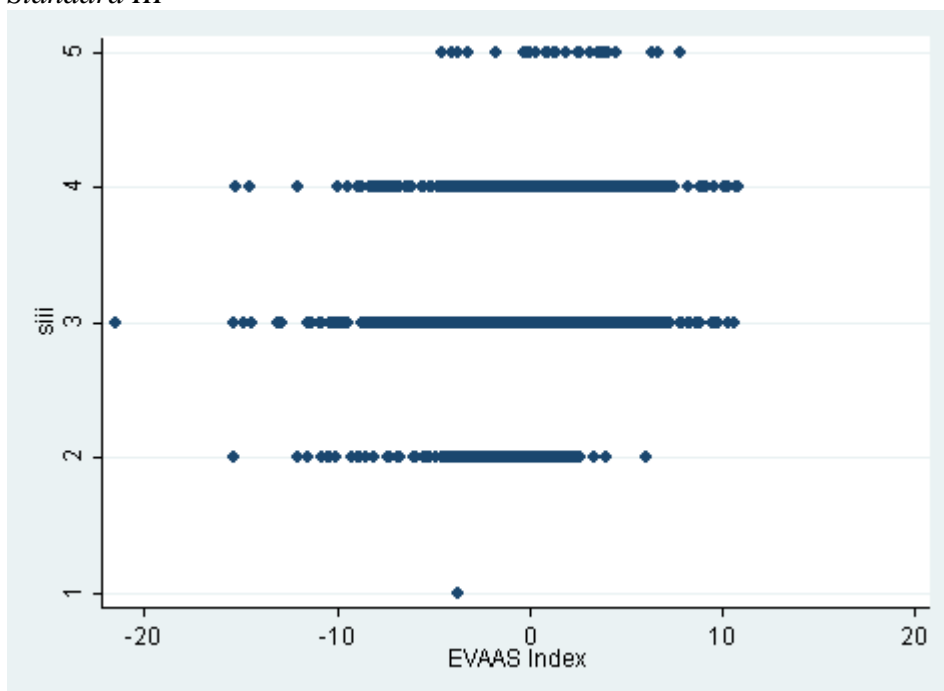
Scatterplot 1
Standard I



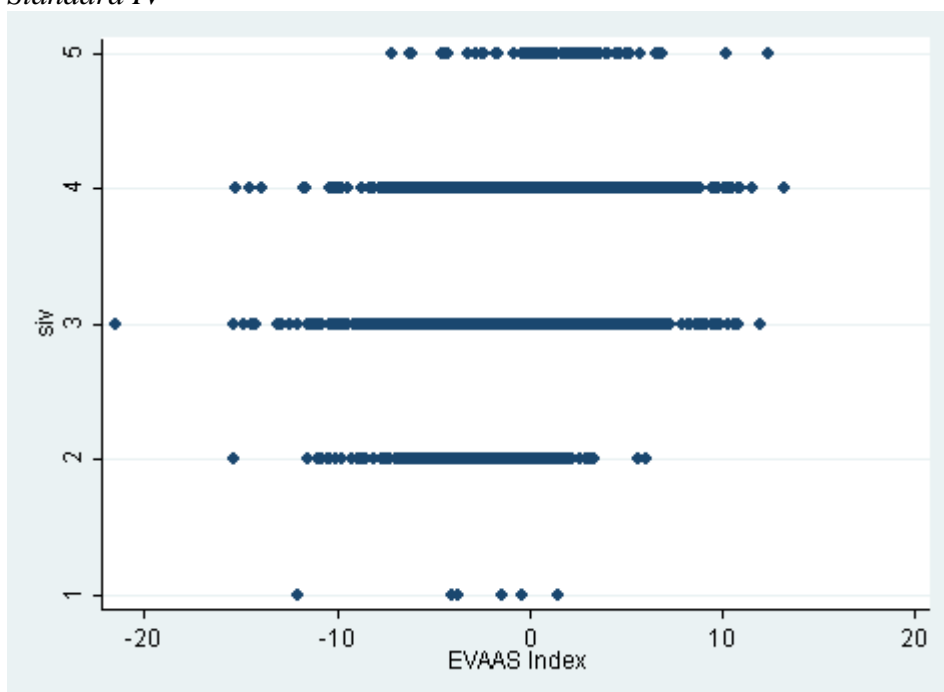
Scatterplot 2
Standard II



Scatterplot 3
Standard III



Scatterplot 4
Standard IV



Scatterplot 5
Standard V

