A QUANTITATIVE STUDY INVESTIGATING

THE EFFICACY BELIEFS OF ELEMENTARY TEACHERS INCLUDING A SPECIFIC

EMPHASIS ON MATHEMATICS

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Educational Leadership

in the

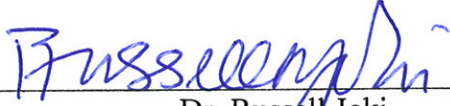Department of Graduate Education

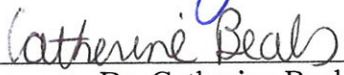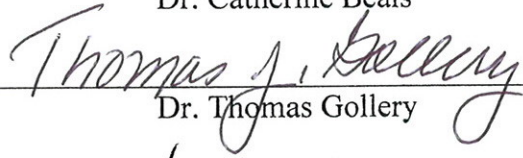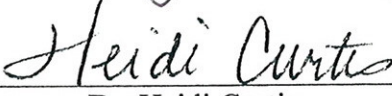Northwest Nazarene University

by

William Fritz

March 2015

Major Professor: Russell A. Joki, EdD

AUTHORIZATION TO SUBMIT DISSERTATION

This dissertation of William Thomas Fritz, submitted for the degree of Doctor of Philosophy in Education with a major in Educational Leadership and titled "A Quantitative Study Investigating the Efficacy Beliefs of Elementary Teachers Including a Specific Emphasis on Mathematics," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies.

Major Professor _____ Date 5-28-15
Dr. Russell Joki

Committee
Members _____ Date 5-28-15
Dr. Catherine Beals

_____ Date 5-26-15
Dr. Thomas Gollery

Department
Administrator _____ Date 4/22/15
Dr. Heidi Curtis

Discipline's
College Dean _____ Date 4/22/15
Dr. Paula Kellerer

ACKNOWLEDGEMENTS

Thank you to my wife, Deanna, for her support and patience during the process of researching and writing this dissertation. I also need to acknowledge my children, Zachary and Joshua, for their understanding during the years I have been pursuing a doctoral degree. Zachary and I have been in school at the same time, and our mutual support should result in earning our degrees in the same month.

I would like to thank the fourth- and fifth-grade teachers who participated in the study for their willingness to answer the survey questions and allow access to their evaluation scores. Although the scores were not individually attributed to participants, it did take trust in the researcher for each subject to participate.

I appreciate the faculty at both Northwest Nazarene University and Seattle Pacific University for their support and guidance. Dr. Lori Werth and Dr. Russell Joki of Northwest Nazarene University and Dr. Richard Smith of Seattle Pacific University have stood beside me the whole way. I also would like to thank Dr. Thomas Gollery, my local statistical adviser, for allowing me to visit regularly and play with numbers. Dr. Cathy Beals has been instrumental in helping me keep this study grounded in the field of mathematics education.

Lastly, I must give credit to my Lord and Savior Jesus Christ for His guidance through this experience. He led me to complete my degree through Northwest Nazarene University, and for that I am thankful. There are times during doctoral study when one wonders whether it will ever be completed, and He has helped me maintain perseverance and self-efficacy to see the *light at the end of the tunnel.*

ABSTRACT

This study analyzes teacher efficacy for fourth- and fifth-grade mathematics teaching and its relationship to teacher and student performance as measured through classroom observation ratings by administrators and mathematical growth of the students. Quantitative correlation methods were used, including point-biserial regression, single regression, and multiple regression. The subjects of the study were 32 teachers in a single school district in Florida. The theoretical framework for the study is rooted in Bandura's construct of self-efficacy as measured by two different instruments: one measuring general teacher self-efficacy and a second measuring mathematics teaching efficacy beliefs. Teacher observation results were based on scores from the Marzano teacher evaluation model, and student growth was measured using Florida value-added mathematics scores for each teacher. Teacher evaluations were based on a combination of observation scores and value-added scores. Significant relationships were identified that validate Bandura's triadic reciprocal framework.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter I

Introduction

Maria, a fourth-grade teacher at Grapefruit Grove Elementary School, has been in the profession for 10 years. She loves her job, receives accolades from parents, is granted gifts for "teacher appreciation week," and her students tend to behave and respond warmly to her teaching. She was named her school's "Teacher of the Year" two years ago by a vote of her colleagues.

In the last few years, though, she has begun to harbor some fear regarding her inadequacies. Under the "old" evaluation system, she had consistently been rated "exemplary" by her principal. The students back then also seemed easier to teach. Her school has become increasingly diverse, and the once suburban neighborhood surrounding the school has become a bit "rougher around the edges." With the downturn in the economy, fewer jobs in the community, and increased competition from charter schools, the school has doubled in its free and reduced lunch rate and is much more diverse than it once was. For the first time this year, her school grade from the state, based on test scores, attendance, and student growth, decreased from an "A" to a "C." It seems harder and harder to reach the students, and her scores on the new Marzano teacher evaluation system note some concerns. The Florida Education Association affiliate in her county has been warning teachers about the possible impacts of state accountability measures, including the right to fire teachers more easily for low evaluation results. Florida is a right-to-work state, and she never felt the need for the union. Last year, however, she joined for the first time out of fear of the unknown. Maria wonders to herself:

- Can I overcome the challenges my students face?
- Why don't parents support the school more?

- Will my evaluations continue to be good?

- Should I believe what other teachers are saying about losing my job due to diminishing evaluation scores?

- Will my teaching be good enough to have most of my students pass state exams, which are predicted to get harder with implementation of the Common Core Standards?

- Will I get the support and professional development I need to succeed?

Maria's reflections and the way she elects to respond to them will reveal much about her belief structure. If she decides that she is capable of overcoming hardship, that she should work hard, that she will persevere, and that her students will succeed because she will find the way to help them achieve, she demonstrates a strong sense of self-efficacy. According to Albert Bandura,

> Perceived self-efficacy is defined as people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Self-efficacy beliefs determine how people feel, think, motivate themselves and behave. (1994, p. 71)

This quantitative study explores the relationship between self-efficacy beliefs and the actual performance of teachers like Maria.

Maria, and a number of her colleague teachers remain hopeful about their ability to reach the children. They help each other by sharing ideas and curriculum materials, and reflecting on progress. The teachers also help one another by offering a listening ear when things are not going well and helping when things get frustrating. Two of her fourth-grade colleagues have had more success than she has had in math gains this year, and they are

coaching her regarding the methods they use. Grapefruit Grove Elementary's sister school, Sawgrass Springs Elementary, meets with teachers from Maria's school during professional development sessions where they share ideas based on what their formative assessment data is telling them. The success of Sawgrass Springs is noteworthy. Maria's principal has also been encouraging, showing confidence in her ability to move forward. Maria also has daily successes in her classroom—"aha" moments that make her day. Bandura's theory discusses these affirming activities by stating that "expectations of personal efficacy are derived from four principal sources of information: performance accomplishments, vicarious experience, verbal persuasion, and physiological states" (1977, p. 191). For Maria to translate knowledge of pedagogy, content knowledge, and situational awareness into effective action, she must have self-efficacy. Bandura states the following:

> Knowledge, transformational operations and constituent skills are necessary but insufficient for accomplished performance. Indeed, people often do not behave optimally even though they know full well what to do. This is because self-referent thought mediates the relationship between knowledge and action. (1986, p. 390)

This study investigates the general self-efficacy beliefs of elementary school teachers like Maria, as well as their self-efficacy in the more specific area of mathematics instruction. Teaching is a human endeavor. As decision makers continue to attempt mathematics reform with the goal of improving students' performance, researchers must look at issues beyond curriculum, assessment, and pedagogy because the beliefs of teachers about whether they are able to make a difference, their self-confidence, their outcome expectancy, and their perseverance are mediating factors in determining the quality of instructional service students receive.

Teacher self-efficacy is the measure of a teacher's confidence in his or her ability to execute appropriate teacher behaviors to positively affect student outcomes (Bandura, 1986). The construct of teacher's sense of efficacy derives from Albert Bandura's general construct of self-efficacy as specifically applied to the educator. Self-efficacy is a key component of Bandura's social cognitive theory, which describes a causal interrelationship among environmental events, personal elements, and behavior (Bandura, 1977, 1986). According to Bandura, knowledge, skills, and the ability to use them will not consistently result in accomplished performances (1986). For success to occur, people must believe that they exercise control over events that affect their lives. In the case of the teacher, one must believe in his or her ability to impact learning. Teachers' sense of efficacy does not, however, stand alone. Teacher experience, knowledge, personal factors, ability to think symbolically, and vicarious experience interrelate with beliefs of efficacy (Bandura, 1989). Self-efficacy affects the teacher's interactions with students in the classroom (Ashton & Webb, 1986; Gibson & Dembo, 1984).

The purpose of this study is to investigate teachers' own beliefs in their ability to affect student achievement, especially in mathematics. The focus is not on teacher mathematical effectiveness, per se. The term *teacher efficacy* can be confused with the notion of teacher effectiveness, so the literature typically uses the terms *teacher's sense of efficacy* or *teacher self-efficacy* (Shaughessy, 2004) to eliminate linguistic and conceptual confusion. This study uses the convention of referring to a teacher's own beliefs as either teacher's sense of efficacy or teacher self-efficacy.

Teacher self-efficacy researchers note two eras, or periods, of study. The first one was from Bandura's seminal study in 1977 to 1998 when Tschannen-Moran, Hoy, and Hoy published their critique of measurement methods for self-efficacy (1998). It is known as the era in which

the construct was conceived and defined and initial research was conducted, but measurement

methods were psychometrically problematic. The second period, after 1998, was when teacher

self-efficacy measurement was stable and the community of scholars addressed research and

measurement gaps from the first era (Klassen, Tze, Betts, & Gordon, 2011; Shaughnessy, 2004).

Improved measurement was primarily due to the development of improved tools for

measurement of teacher self-efficacy, refined most notably at Ohio State University (Tschannen-

Moran & Hoy, 2001).

The relationship between teacher self-efficacy and elementary mathematics instruction is

of particular importance because elementary teachers in the United States tend to be generalists

with less content-specific college coursework than their secondary colleagues, and because the

instruction elementary students receive in mathematics builds the foundation for their skills,

abilities, and efficacy belief structures around mathematics (Rowland, Huckstep, & Thwaites,

2005; Stigler & Hiebert, 2009).

Statement of the Problem

Concern about the state of mathematics instruction in the United States is not a new

phenomenon. In fact, documented efforts to reform math instruction date back to the late 19th

century. Mathematics education reform has endured post-Sputnik reforms, new math, A Nation at

Risk, NCTM Standards, No Child Left Behind, and the Common Core Standards, with tension

between those who advocate for rote instruction and others who advocate for a more exploratory

curriculum being the pervasive theme (Abbott, Baker, Smith, & Trzyna, 2010; Kanold & Larson,

2012; Mewborn, 2013; Nisbet & Warren, 2013; Wilson, 2013). Researchers continue to conduct

comparisons of American mathematics students and those of the world, with the Trends in

International Mathematics and Science Studies series being one persistent example (Provasnik,

Gonzales, & Miller, 2009). Recently, U.S. President Obama announced an initiative and funding to improve the country's mathematics instruction by training teachers and researching models for science, technology, engineering, and mathematics (STEM) programs (Larson, 2012).

Attempts to improve mathematics instruction frequently focus on the curriculum, assessment methods, and instructional pedagogy (Waters, Marzano, & McNulty, 2003). A typical example of a recommended strategy for improving mathematics instruction comes from the Trends in Mathematics and Science Study reports (Provasnik et al., 2012). TIMSS is a widely referenced series of video-based studies that compares international mathematics instructional methods and results. This recommended strategy includes a focus on the kinds of mathematics students encounter, methods for developing concepts and procedures, tasks students are expected to complete, the teacher's role, lesson organization, and attitudes toward reform (Stigler & Hiebert, 1997, 2009; Provasnik et al., 2012). More recently, with the advent of the Common Core State Standards for Mathematics (CCSSM), strategies include focusing on curriculum, instruction, assessment, interventions, and professional development (Larson, 2012), with continued neglect for the self-reflective efficacy beliefs of teachers.

Thames and Ball (2013) recall the persistent criticism of mathematics education for over 50 years, and propose an agenda to move forward. Included within this agenda are a coherent mathematics curriculum, a supportive learning environment, proper educational infrastructure, and skilled teaching, and self-efficacy beliefs of teachers are implicit in some of the proposal. When discussing teacher preparation, the writers discuss teachers' feelings of safety, which relate to Bandura's personal factor element. They also discuss performance measures within teacher preparation programs, which are similar to Bandura's suggestion of the need for mastery experiences. Also suggested is a support program for new mathematics teachers that provides

vicarious experiences to assist teachers. Although Thames and Ball allude to the need for self-efficacy, the construct is not explicit in their writing.

An explicit construct is important because teacher self-efficacy mediates teachers' thoughts, perseverance, and action. Bandura reminds us of the critical role of self-efficacy beliefs:

> Among the types of thoughts that affect action, none is more central or pervasive than people's judgments of their capabilities to deal effectively with different realities. It is partly on the basis of self-percepts of efficacy that they choose what to do, how much effort to invest in activities, how long to persevere in the face of disappointing results, and whether tasks are approached anxiously or self-assuredly. (1986, p. 21)

Reformers must focus on the beliefs of the teachers, as each of the recommended elements are mediated through the beliefs of the educator; unfortunately, research regarding the belief structures for the primary deliverers of mathematics instruction remain neglected in the literature (Klassen et al., 2011; Holzberger, Philipp, & Kunter, 2013).

Further analysis of the sources of teachers' sense of efficacy for instruction is needed. This study emphasizes self-efficacy and general teacher performance. Although the study investigates overall performance of teachers as it relates to efficacy, a specific area of focus is how mathematics instruction is related to self-efficacy, an area further neglected in the literature. The study also identifies the proportions of influence self-efficacy, experience, professional development, and mathematics coursework have on teacher evaluation performance.

Background

The theoretical construct on which the study is based is Bandura's theory of teacher self-efficacy. Self-efficacy is the measure of one's belief in his or her own ability to affect results (Bandura, 1977, 1986, 1997). More specifically, teacher self-efficacy is the educator's belief in his or her ability to impact achievement or behavior of students (Tschannen-Moran, Woolfolk-Hoy, & Hoy, 1998). The theory involves three factors that are interdependent and are described as having triadic determination: behavior, environmental factors, and psychological factors (Bandura 1997). Sources of self-efficacy, as theorized by Bandura, are mastery experiences, vicarious experiences, verbal persuasion, and psychological factors (Bandura, 1997). Although mastery experiences are typically most powerful, each of these sources of efficacy is believed to impact one's efficacy levels and eventual behavior.

Self-efficacy is closely related to Rotter's earlier theory of locus of control (1966). Self-efficacy is a component attributable to the individual, but later theorists have expanded the concept to describe groups' beliefs, called *collective efficacy*. In education, self-efficacy is also an element of a broader construct entitled *academic optimism*, which combines self-efficacy beliefs, academic rigor, and trust into an overarching construct (Hoy, Tarter, & Hoy, 2006).

If a connection exists between teacher self-efficacy and a teacher's ability to help students achieve, it makes sense that a teacher's self-efficacy would also be connected to that teacher's evaluation. Therefore, this study also explores this connection. Teacher evaluation in the state of Florida was changed legislatively in 2011 (Student Success Act) to require enhanced teacher evaluation systems and a value-added element. The Department of Education, consistent

with Florida statutes, adopted the Marzano teacher evaluation system (Marzano, 2011). This system involves four domains (Marzano, 2011):

- Classroom strategies and behaviors

- Planning and preparing

- Reflecting on teaching

- Collegiality and professionalism

Within each domain are specific high-yield practices denoted as dominant elements and scored. Examples of the 60 elements include the following (Marzano, 2013):

- Providing clear learning goals and scales (rubrics)

- Establishing classroom routines

- Examining similarities and differences

- Organizing students for cognitively complex tasks

- Noticing when students are not engaged

- Managing response rates

- Demonstrating value and respect for low expectancy students

- Using available technology

- Evaluating the effects of individual lessons and units

- Seeking mentorship for areas of need or interest

The system provides a framework for principals to evaluate teachers. Final evaluations, however, also involve a value-added score, which is derived by the State of Florida using test data in mathematics and reading. The value-added score is an indicator of how a teacher's students performed when compared with similar students' predicted scores in the state. A positive score means the students exceeded what was predicted and a negative score is an

indicator that the students performed less well than predicted (Braun, 2005; Florida Department of Education, 2012).

The Florida evaluation model, with the combined metrics of Marzano evaluation scores and value-added scores, has received much scrutiny from Florida teachers (Baeder, 2011; O'Keefe, 2012). Teachers describe the new model as "artificial, frustrating, (and) humiliating" (Postal, 2012). In 2013, the Florida Education Association financed a lawsuit against the State Department of Education and Commissioner of Education over the evaluation model (Cook v. Stewart, 2014). The Department of Education, however, has prevailed. The story is typical of those publicized by the Florida Education Association. The named plaintiff was previously rated "exemplary" in her evaluations and was named "teacher of the year" for her county. Under the new model, she was rated "Needs Improvement" in her evaluation, which causes her to forfeit merit pay and is the first step toward just cause for termination.

The model is too new, however, to have received much psychometric scrutiny, and the research behind the model is predominantly self-referent (Haystead & Marzano, 2009; Marzano, 2013; Marzano, Toth, & Schooling, 2012). Because the model is new, this study adds to the body of literature that is sure to evolve given the high-stakes nature of evaluations under the Race to the Top federal funding structure (U.S. Department of Education, 2009). Additionally, Marzano's research has focused on how the practices impact student achievement (Haystead & Marzano, 2009; Marzano, 2013; Marzano et al., 2012). Marzano's evaluation model has been the independent variable with student performance as the dependent variable. This study used Marzano evaluation scores as the dependent variable, adding a converse dimension to existing research.

Over the past 10 years, value-added modeling has been increasingly utilized to evaluate teachers and determine types of merit pay in states (Stewart, 2006). Although some have been critical of using high-stakes models to make high-stakes decisions (Braun, 2005; Wainer, 2011), others have expressed balanced optimism regarding the model (McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Koedel & Betts, 2007; Stewart, 2006). This study is one of the first to analyze value-added scores of teachers by using a sample that includes teachers who work within a model that is wide spread and includes an accountability system.

Research Questions

Three research questions are addressed. Each lends itself to correlation methodology but through different elements of teacher performance, including observations by administrators, student growth scores in mathematics, and overall evaluation results of teachers. The central research questions addressed are:

1. What is the relationship between teachers' levels of self-efficacy and teacher observation scores derived from the Marzano evaluation system?

2. What is the relationship between teachers' sense of self-efficacy and their Florida State value-added scores in mathematics?

3. What proportion of influence do the following have on a teacher's evaluation rating: teacher sense of efficacy, as measured by the Tschannen-Hoy instrument; mathematics teaching efficacy, as measured by the MTEBI; years of teaching experience; mathematics professional development level; and mathematics coursework attained?

Description of Terms

Terms specific to the study relate to the construct of self-efficacy and measurements of teacher and student performance. Self-efficacy measures were the independent variables for the study, and the teacher and student performance measures were the dependent variables.

Instructional performance. A dependent variable, measured by principals through use of the Marzano evaluation model (Marzano, 2013).

Locus of control. The extent to which individuals believe they can control events that affect them (Rotter, 1966).

Marzano score. A dependent variable representing the overall teacher evaluation rating by the administrator based on Robert Marzano's teacher evaluation system (Marzano, 2013). Scoring levels ranked from most effective to least effective are innovating, applying, beginning, and not using.

Mathematics professional development level. The number of professional development hours in which each subject has participated during the last five years. This element is self-reported by each subject of the study.

Teacher self-efficacy in mathematics A dependent variable representing the measure of a teacher's confidence in his or her ability to positively affect student outcomes in mathematics, as measured by the Mathematics Teaching Efficacy Beliefs Instrument (Enochs, Smith, & Huinker, 2000).

Teacher self-efficacy. An independent variable representing the measure of a teacher's confidence in his or her ability to execute appropriate teacher behaviors to positively affect student outcomes, as measured by the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Hoy, 2001)

Teacher's sense of efficacy. This term is used interchangeably and is synonymous with the term teacher self-efficacy (Shaughnessy, 2004).

Value-added measure (VAM). A dependent variable. The State of Florida derives a teacher's VAM score by analyzing his or her students' composite growth in the mathematics results on the Florida Comprehensive Achievement Test (FCAT) and comparing it with the expected growth of similar sets of students state-wide.

Significance of the Study

This study increases the body of literature regarding teachers' sense of efficacy and elementary mathematics instruction. There is a gap in the professional literature related to self-efficacy beliefs of elementary school teachers in the area of mathematics instruction (Klassen et al., 2011; Shaughnessy, 2004; Wyatt, 2012). Findings will assist administrators and practitioners with motivation and preparation toward the goal of improved mathematics instructional delivery. A large portion of the research conducted in the area of teacher self-efficacy is based on qualitative methodology, and this study will add to the body of quantitative research. Because the study investigates the relationship between teacher self-efficacy and student achievement, the results will have implications for the teaching profession, given that the primary purpose of schooling is academic preparation.

Other unique attributes to this study are use of value-added data and use of Marzano evaluation data. Legislation regarding use of student growth data is a recent phenomenon in education. In Florida, legislation instituting value-added attribution to teachers began in 2011 (Florida Department of Education, 2012) with Senate Bill 736 (Student Success Act, 2011). This law is in response to Florida's Race to the Top federal grant application, and similar models are being developed in a number of other states (Braun, 2005; Wainer, 2011). With the advent of

this new system, studies have been conducted regarding the psychometrics of the model, but not using value-added scores from Florida as a dependent variable; the model is simply too new.

Another result of federal requirements is implementation of new evaluation systems in the United States. The Marzano system of evaluation (Marzano, 2013) contains 60 elements on which teachers can be evaluated. These elements describe strategies that have been identified to predict improved student performance if implemented by teachers (Haystead & Marzano, 2009). The field testing of the evaluation model ended in 2009, and the system is being implemented now in school systems throughout the country. To date, studies involving value-added models are rare, and there are no published studies where Marzano ratings are a dependent variable. There are, however, a number of studies in which implementation of the strategies is the predictor variable, and they attempt to link the Marzano traits to improved student learning (Haystead & Marzano, 2009; Marzano et al., 2012). There is one study that connects teachers' sense of efficacy with results on the Danielson evaluation model, which is similar to the Marzano model (Heneman, Kimball, & Milanowski, 2006). The Danielson study used a correlation model and demonstrated a positive and significant relationship between teacher self-efficacy and evaluation results.

Kleinsasser (2014) recently reviewed 30 years of self-efficacy literature, including 111 articles published in *Teaching and Teacher Education.* He concludes that there remains a need for research that creates practical meaning from the research. Kleinsasser also notes that the research continues to confirm the complexity, specificity, and situational variability of teacher self-efficacy. Given the recent introduction of teacher observation systems like Marzano's, value-added systems, and accountability measures, this study attempts to answer Kleinsasser's call.

This study extends the knowledge base regarding the relationship between teacher self-efficacy and student learning, but it also breaks new ground by using value-added scores and Marzano evaluation results as dependent variables. The other key element of this study is identification of whether a relationship exists between predictors, including teacher self-efficacy levels, and evaluation results. By use of a multiple regression model, the study identifies the degree of predictive value teacher self-efficacy has on teacher evaluations.

Overview of Research Methods

The sample involved teachers from a school district in the state of Florida. Fourth- and fifth-grade teachers were chosen due to the availability of value-added mathematics scores for these teachers. The school district chosen includes urban, suburban, and rural regions, and the teacher population reflects the demographics of the overall population in the state of Florida (see Appendix A).

Quantitative research methods were used to answer the research questions. The first research question analyzed the correlation between teacher self-efficacy and evaluation results using the Marzano evaluation model. The evaluation question used a point-biserial correlation. Correlational methods were also used to answer the second research question, the relationship between measures of teacher self-efficacy and value-added scores. For this question, a single regression was used. The third question analyzed the degree of relationship between general teacher self-efficacy, mathematics teaching self-efficacy, mathematics-specific professional development, college mathematics coursework, and years of experience on evaluation ratings. For the third question, a multiple regression analysis was utilized, because the independent factors were either scale or ordinal variables and the dependent measure was a scale variable.

Levels of teacher self-efficacy were measured using both the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Hoy, 2001), and the Mathematics Teaching Efficacy Beliefs Instrument (Enochs, et al., 2000). The measurement tools are generally accepted to be valid and reliable measures of self-efficacy, and the dual use of these tools, which have slightly different conceptual foundations and levels of specificity, enhanced the richness and scope of the study (Klassen et al., 2011).

Chapter II

The Literature Review

Introduction

The conceptual framework for the study is Bandura's construct of teacher self-efficacy (Bandura, 1977, 1986, 1997). The literature regarding this construct includes explorations of the history and evolution of teacher self-efficacy, how self-efficacy is related to Bandura's more general social-cognitive theory (1986), manners of differentiating between self-efficacy and self-concept, and how researchers measure self-efficacy for teachers.

Research regarding teacher self-efficacy is generally believed to have evolved during two eras (Klassen et al., 2011). During the first era, from 1977 to 1998, measurement had not been fully developed and there were gaps in much of the research. After 1998, measurement tools were believed to be more reflective of the actual construct of teacher self-efficacy as distinct from locus of control (Henson, 2001; Tschannen-Moran & Hoy, 2001). During this second era, not only did the research have more stability, but self-efficacy was also identified as an essential element in the broader constructs of academic optimism (Hoy et al., 2006 and collective efficacy (Eells, 2011).

This review outlines specific predictors of teacher self-efficacy, how self-efficacy is connected to mathematics learning levels for students, and how self-efficacy is connected with a variety of classroom practices of teachers. The literature review describes value-added modeling, recognizes questions regarding its use, and identifies a lack of studies validating whether a relationship exists between self-efficacy of teachers and their evaluation scores in the modern era of value-added scores and enhanced evaluation systems. Each of these elements was explored within this study.

Conceptual Framework

      Teacher self-efficacy as the central theme. This study is rooted in Bandura's theory of self-efficacy, specifically as it pertains to elementary school teachers. Teacher self-efficacy is the measure of a teacher's confidence in his or her ability to execute appropriate actions to positively affect student outcomes (Bandura, 1986). The construct of teacher efficacy is a teacher-specific example of Albert Bandura's more general construct of self-efficacy. Self-efficacy is a key component of Bandura's social cognitive theory, which describes a causal interrelationship among environmental events, physiological elements, and behavior (Bandura, 1977, 1986). Bandura writes specifically regarding teachers' perceived efficacy in his books *Self-Efficacy in Changing Societies* (1995) and *Self-Efficacy: The Exercise of Control* (1997). According to Bandura, knowledge, skills, and the ability to use them are insufficient for accomplished performances (1986). For success to occur, people must believe that they can control the events that affect their lives. To perform successfully, teachers must believe in their ability to impact learning. Teachers' sense of efficacy does not, however, stand alone. Teacher experience, knowledge, personal factors, ability to think symbolically, and vicarious experience interrelate with beliefs of efficacy (Bandura, 1989). Because self-efficacy affects the way teachers feel, think, and behave (Bandura, 1995), self-efficacy affects the teacher's interactions with students in the classroom.

      The relationship between teacher self-efficacy and elementary mathematics instruction is of specific importance because elementary students' performance in mathematics tends to lag behind performance in other content areas. The instruction elementary students receive in mathematics builds the foundation for their future mathematical study, and teacher beliefs mitigate instructional effectiveness (Rowland et al., 2005; Stigler & Hiebert, 2009).

Theoretical underpinnings. Within this study, teacher self-efficacy is explored by identifying correlations with other theories and topics, which act as underpinnings for the overall theoretical framework. Bandura's theory of teacher self-efficacy (1977), the specific application of teacher self-efficacy to the field of mathematics, Marzano's teacher observation scores, and value-added mathematics scores each act as variables within this study. The theoretical framework attempts to identify relationships between and amongst these elements.

Although not primary components of this study, other elements are included to identify their degree of influence on teacher evaluation ratings as either mediators or as they enhance the relationship between efficacy beliefs and teacher performance. Within the study design, these are not considered primary variables but are necessarily incidental to the framework. These include years of teaching, college-level coursework successfully completed in mathematics, and the quantity of professional development in mathematics teaching completed by each teacher.

Relationship Between Teacher Self-Efficacy and Other Theories

Bandura's article, "Self-efficacy: Toward a unifying theory of behavioral change" (1977), introduces the concept of self-efficacy. In the article, he describes sources of self-efficacy: performance accomplishments, vicarious experiences, verbal persuasion, and emotional arousal. The theoretical underpinnings are attributed, almost exclusively, to his 1960s and 1970s contemporaries in psychology. Twelve of Bandura's citations are other articles written by Bandura. That being said, it is helpful to view the construct of self-efficacy by looking at self-efficacy's relationship with locus of control theory (Rotter, 1966), considering its role within the more global construct of social cognitive theory (Bandura, 1986), and comparing and contrasting self-efficacy with self-concept.

Locus of control and the RAND studies. Early work regarding self-efficacy was based on Rotter's (1966) locus of control theory. Two studies conducted by the RAND Corporation are acknowledged as the birth of the teacher self-efficacy construct. RAND is a not-for-profit policy and decision-making research corporation whose researchers cite Rotter's 1966 article entitled "Generalized Expectancies for Internal Versus External Control of Reinforcement" as the inspiration for the studies in the mid-1970s. Within Rotter's construct, individuals attribute success or failure to either themselves (internal locus of control) or circumstances beyond their own control, such as actions of others or events unrelated to themselves (external locus of control). The Armor et al. (1976) study was the first that identified a relationship between reading performance of minority students and beliefs represented by the following two survey items: (a) "When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment," and (b) "If I try really hard, I can get through to even the most difficult or unmotivated students" (p. 2). The first belief came to be identified as general teaching efficacy (GTE) and the second as personal teaching efficacy (PTE).

The second RAND study (Berman, McLaughlin, Bass, Pauly, & Zellman, 1977) found teacher-efficacy to predict continuation of federally funded projects beyond the end of project funding. GTE and PTE were the commonly used attributes of teachers' sense of efficacy in research until Tschannen-Moran and Hoy's Teachers' Sense of Efficacy Scale was developed in 2001 (Gibson & Dembo, 1984; Ashton & Webb, 1986; Henson, 2001; Tschannen-Moran & Hoy, 2001).

The theories of locus of control and self-efficacy are so closely related that much of the literature related to measurement of efficacy into the 2000s attempted to psychometrically ensure

distinction between the constructs (Denzine, Cooney & McKenzie, 2005; Henson, 2001; Henson, Kogan, & Vacha-Hasse, 2001; Hoy, 2000). Certainly, the concept of locus of control is embedded in self-efficacy. However, even a person with strong internal locus of control can have his or her sense of efficacy altered due to lack of success, negative modeling, information (vicarious) that leads to doubt, or excessive self-imposed or superimposed expectations (Bandura, 1986). Bandura (1986) also suggests that there is an element beyond efficacy: outcome expectancy. Outcome expectancy is the vision of the likely results of one's behavior. Some have argued a causal link between outcome expectancy and self-efficacy; however, Bandura disagrees with a pure cause-effect relationship between expected outcomes and self-efficacy beliefs (Williams, 2010). Behavior is closely linked with both beliefs and projected outcomes.

Teacher self-efficacy and social cognitive theory. Social cognitive theory describes the triadic reciprocal relationship among environment, behavior, and internal processes (see Figure 1) (Bandura, 1977, 1986, 1991; Henson, 2001). Environmental factors include actual experiences of the individual as well as vicarious ones. Behavior includes one's actions. The internal processes include cognitive, affective, and biological processes. One's thoughts and actions are not simply dictated by environment or biology, but rather are part of an ongoing interplay between the world and one's thoughts and behavior. In Bandura's original conception of self-efficacy, he described it as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (1977, p. 3). Study of self-efficacy beyond his initial conception focused on the interplay among self-efficacy, the environment, and behavior. Experiences certainly have an impact on one's self-efficacy, and the level of self-efficacy impacts behavior. Bandura (1986) argues that beliefs have a stronger influence on behavior than does environment.

Figure 1

Triadic Reciprocal Determinism

Personal Factors –
Biological, Cognitive and Affective

Behavioral Factors –   Environmental Factors –
What we do   What occurs around us
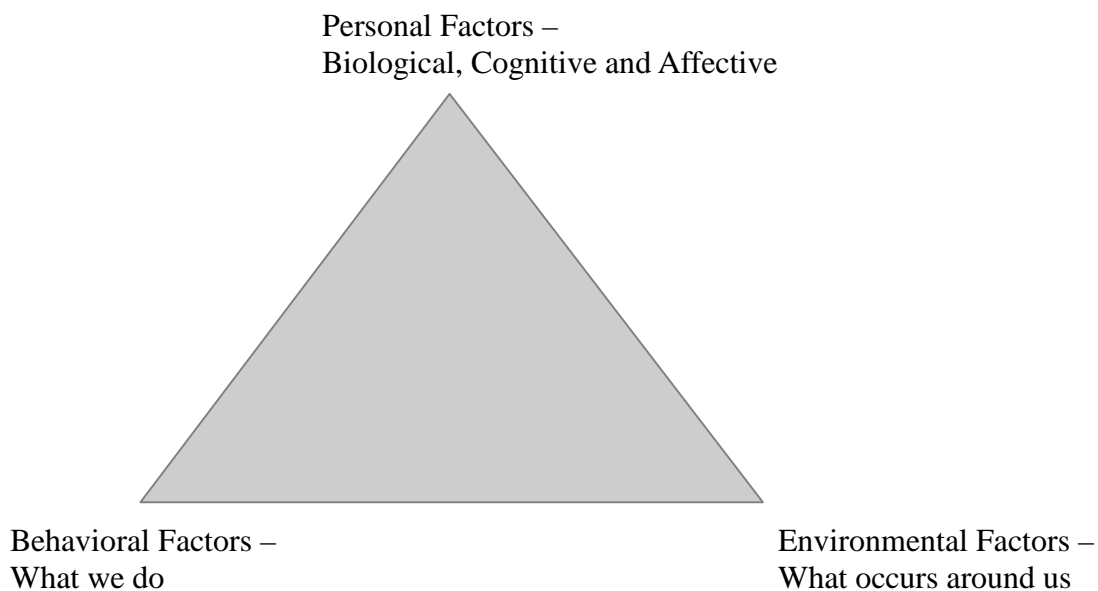
Note. Reciprocal determinism is a construct in which action, personal factors, and the environment all impact one another. Self-efficacy is a key personal factor that mediates behavior and how people relate with their environment. Included with permission. See Appendix B. (Bandura, 1986, p. 24)

In a 2004 retrospective article, Bandura reflects on the cold reception he received from the psychological community when he introduced social cognitive theory (Bandura, 2004). The predominant schema of the 1950s and 1960s remained that of behaviorism, and Bandura proposed that cognition was a mediating factor to conditioning. He asserted that, "People of high perceived self-efficacy set motivating goals for themselves, maintain commitment to them, expect their efforts to produce favorable results, view obstacles as surmountable and figure out ways to overcome them" (p. 623). He reminded the psychological community that human agency is a key element that must be contemplated when looking at environmental influences and behavior. Related to low self-efficacy, he asserted that "people … are easily convinced of the futility of effort in the face of impediments" (p. 623). Social cognitive theory is now the accepted schema, but the transition to widespread acceptance is a recent phenomenon.

The theory of triadic reciprocal determinism has proven difficult to verify quantitatively. Measuring causal effects is difficult in education due to lack of ability to implement random assignment of subjects, but Bandura's model is especially challenging due to its reciprocal nature. Each cause is also an effect, and the converse is also theorized to be true. An additional challenge is that the theory is not necessarily longitudinal in nature; causes and effects can have a simultaneous effect, or there may be a state of equilibrium that evolves. One study by Williams and Williams (2010) does verify the reciprocal determinism theory in a study with a large sample size that takes place in 30 countries. The approach isolated each variable to identify the direction of influence of each variable on the other factors in Bandura's model; as a result, they were able to suggest a causal effect of behavior on beliefs, as well as the converse. The international nature of this study also shows that the triadic reciprocal determination model is not necessarily prone to cultural bias (Williams & Williams, 2010).

According to Bandura (1977), four information sources impact an individual's sense of personal efficacy:

- performance experiences: a person's personal successes and failures,

- models: observed successes and failures of others,

- verbal persuasion: messages the individual receives from others, and

- physical or emotional reactions: reinforcements or deterrents that impact future decisions.

Although teacher self-efficacy is only a single element in the interdependent social cognitive relationship as related to learning in schools, it is central to the possibility of success. The filter of teacher self-efficacy affects all interactions between teacher and student (Bandura, 1986). This makes sense, because the teacher's sense of efficacy has been found to correlate positively with the teacher's level of expectations, level of effort, affective elements within the

classroom, classroom management approaches, curriculum choice, the way the teacher communicates, levels of job stress, teacher engagement, levels of teacher emotional exhaustion, and instructional methods (Ashton & Webb, 1986; Gibson and Dembo, 1984; Multon, Brown, & Lent, 1991; Reilly, Dhingra, & Boduszek, 2014; Skaalvik & Skaalvik, 2014). According to Tschannen-Moran and Hoy (1998), sources of efficacy information (verbal persuasion, vicarious experiences, physiological arousal, and mastery experiences) combine with analysis (of the teaching task and personal competence) to create a level of efficacy.

This level of efficacy, in turn, affects how teachers deliver instruction to students. On a theoretical level, Bandura (1991) asserts the following:

> People's beliefs in their efficacy influence the choices they make, their aspirations, how much effort they mobilize in a given endeavor, how long they persevere in the face of difficulties and setbacks, whether their thought patterns are self-hindering or self-aiding, the amount of stress they experience in coping with taxing environmental demands, and their vulnerability to depression. (p. 257)

More specific attributes associated with teachers who possess strong levels of efficacy are discussed later.

It should be noted that teacher self-efficacy does not exist as a static construct. According to Bandura's model (1986), there exists constant interaction between behavior, environment, and personal factors (including levels of efficacy). Bandura (1997) contends that people's level of motivation, affective states, and actions are based more on what they believe than on what is actually true. For that reason, success can improve one's self-efficacy, and improved efficacy can increase healthy motivation and effective behavior. Likewise, failure can lead to decreased self-efficacy, and low efficacy can cause one to have self-doubt and inactivity (lack of behavior) or

even self-destructive behavior. Self-efficacy, therefore, is not static and is both a cause and an effect within social cognitive theory.

Comparison and contrast between self-efficacy and self-concept. Although self-efficacy and self-concept are both forms of self-belief, the constructs are different from one another. Self-concept is a more complicated construct, including both cognitive and affective beliefs toward the self, and self-efficacy concerns cognitive judgments of one's ability based on mastery criteria (Bong & Clark, 1999). Critics have noted that self-concept is not necessarily a well-defined or stable construct (Marsh, Walker, & Debus, 1990).

Meta-analyses of the two constructs have been conducted. Self-efficacy has been found to have a direct and mediating effect on student learning (Multon et al, 1991). On the other hand, a relationship between self-concept and student performance, as demonstrated in a meta-analysis of 128 studies, is unclear (Hansford & Hattie, 1982). These meta-analyses do contain different characteristics and methodologies from one another but suggest that the constructs have different connections when considering achievement.

The operational definitions of self-concept and self-efficacy differ in that self-concept is less clearly defined than self-efficacy (Hansford & Hattie, 1982). A large number of studies regarding self-concept do not even furnish explicit definitions (Bong & Clark, 1999; Byrne, 1984; Marsh, 1990). Self-efficacy scales, including those of Tschannen-Moran and Hoy (2001) and Gibson and Dembo (1984), have generally stayed close to the original definition proposed by Bandura (1977).

In Bong and Clark's analysis (1999), the authors assert that self-concept researchers typically utilize a global approach, and self-efficacy researchers study with tools that measure self-efficacy in specific situations. The other difference reported by Bong and Clark is that self-

efficacy research typically measures achievement with a specific task, but self-concept research uses more global indexes, like standardized test scores or grade point average (1999). Not surprisingly, Bandura (1997) advises that self-efficacy beliefs may vary greatly if the task is specific in its content, complexity, or standards.

Collective efficacy. Although it is not the primary topic of this study, a mention of collective efficacy should be offered so as to clarify the construct and ensure understanding of the scope of this particular study. Self-efficacy is a construct that is specific to an individual and is highly personal. Several researchers, however, have expanded the construct of self-efficacy to describe the self-efficacy of a group. In the education realm, collective efficacy describes the overall sense of teaching self-efficacy of the staff. Bandura first described the collective construct in 1986, by stating the following:

> The strength of groups, organizations, and even nations lies partly in people's sense of collective efficacy that they can solve their problems and improve their lives through concerted effort. Perceived collective efficacy will influence what people choose to do as a group, how much effort they put into it, and their staying power when group efforts fail to produce results. (p. 449)

Several researchers have shown collective efficacy in schools to have a stronger connection with school-wide achievement than teaching self-efficacy limited to a particular teacher (Bandura, 2000; Eells, 2011; Klassen et al., 2011; Mills, 2009).

Academic optimism. Hoy (2005) also has focused on the construct of collective efficacy, identifying it along with academic emphasis and a trusting environment as key elements in a broader construct called academic optimism. Emerging research has shown a positive correlation between higher levels of academic optimism and overall strong achievement levels of schools

(Hoy, Tarter, Woolfolk-Hoy, 2006; Hoy, 2012). Much like the triadic determinism concept of Bandura (1997), Hoy (2005) has described academic optimism as having three elements, of which collective efficacy is one, that are essential and that possess a triadic determination relationship. More recently, Boonen, Pinxten, Van Damme, and Onghena (2014) have verified a positive correlation between academic optimism and both mathematics achievement and reading comprehension for primary school students, overcoming socioeconomic predictors. This study involved 1,375 staff members and 3,538 fifth-grade students within 117 schools.

Researching and Measuring a Teacher's Sense of Efficacy

Inherent challenges exist when researching teacher self-efficacy. One challenge within social cognitive theory is that any cause is also an effect given Bandura's inter-relational approach (Bandura, 1986). Ongoing questioning exists regarding the definition of teacher self-efficacy as a discrete construct; it is often similar to Rotter's concept of locus of control (1966), and different studies elicit different operational characteristics (Henson, 2001). Measures of self-efficacy can vary moment by moment and task by task, because the environment affects them. The stability of measurements, therefore, can be a challenge (Bandura, 1986, 1997). Because a tenet of self-efficacy research is that measures of academic performance are usually task specific (rather than general), problems exist with implementation of proper controls, as many studies are conducted with real teachers in real schools (Bong & Clark, 1999).

Prior to 2000, researchers utilized a two-factor assessment to measure the construct of teacher self-efficacy (Denzine et al., 2005; Gibson & Dembo, 1984; Tschannen-Moran & Hoy, 2001). The tools commonly used prior to 2001 were Gibson and Dembo's Teacher Efficacy Scale (TES) (1984) and Bandura's Instructional Efficacy Scale (1997). The two elements included were general teaching efficacy (GTE) and personal teaching efficacy (PTE). General teaching

efficacy is described as the teacher's outcome expectations—whether the teacher believes the children can succeed. Personal efficacy is the teacher's belief in his or her personal ability to influence student learning—whether the teacher has the ability to reach the students (Gibson & Dembo, 1984).

The Gibson and Dembo tool, most commonly used prior to 2001, was found to have concerns related to discriminant validity for GTE and PTE scores when viewed in light of other assessments (Henson, 2001; Tschannen-Moran et al., 1998). Guskey and Passaro (1994) had previously reported that PTE and GTE factors actually correspond to internal versus external orientation. This is not surprising, given that the source of Gibson and Dembo's TES could trace its origins to locus of control theory (Rotter, 1966).

In 2001, Woolfolk, Tschannen-Moran, and Goddard developed the Teachers' Sense of Efficacy Scale (TSES), also referred to as the Ohio State University tool, which measures an individual teacher's efficacy and a teacher's sense of collective efficacy (Tschannen-Moran & Hoy, 2001). This tool addresses the psychometric critique that had plagued the earlier tools (Henson, 2001; Shaughnessy, 2004). The new tool uses elements from the Gibson and Dembo tool but includes elements consistent with Bandura's 1997 conceptualization (Woolfolk & Hoy, 2001). Additionally, the tool accounts for the specific situational nature of teacher self-efficacy; efficacy differs depending on such elements as resource allocation, class size, and subject matter (Woolfolk & Hoy, 2001). A 2006 study (Heneman et al.) verified the construct validity of the Ohio State University tool within a study of 180 elementary school teachers in Nevada, saying that its psychometric features should make it the preferred measure for a teacher's sense of efficacy. Klassen et al. (2011) conducted an analysis of all peer-reviewed articles published between 1998 and 2009, and findings indicated strong preference for the Ohio State University

tool's psychometric attributes and discounted virtually all other tools. A study by the Wisconsin Center for Education Research further confirmed the psychometric properties of the Teachers' Sense of Efficacy Scale (Heneman et al., 2006). A 2013 study (Vieluf, Kunter, & van de Vijver) verified the same factor structure as generalizable to international teachers.

Specific to mathematics, Enochs, Smith, and Huinker (2000) developed the Mathematics Teaching Efficacy Beliefs Instrument. This tool was closely related to an earlier science instrument by the same researchers. Their analysis, with a sample of 324 pre-service elementary teacher education students at six universities in four states, confirmed two independent factors: personal mathematics teaching efficacy (PMTE) and mathematics teaching outcome expectancy (MTOE). In a 2011 analysis of the psychometric properties of the Enochs tool, a Turkish adaptation of the tool demonstrated the same two factors, suggesting strong internal consistency. The sample size of the 2011 study, 1,355 in-service teachers in 368 schools, adds to the strength of the results. The literature review within the 2011 report refers to nine other published analyses of the Enochs tool that verify its validity and reliability. On the other hand, Klassen et al. (2011) discount the validity of the Enochs tool, indicating it is flawed because its basic structure is aligned with the discounted Gibson and Dembo (1984) tool.

Teacher Self-Efficacy and its Relationship with Student Learning

The study of teacher-efficacy began with the RAND studies (Armor et al., 1976; Berman et al., 1977) and the research of Ashton and Webb (1986) linking general teaching efficacy (GTE) and personal teaching efficacy (PTE) with math achievement and language achievement on the Metropolitan Achievement Test, respectively. The Ashton and Webb study should be met with some skepticism due to its limited sample of at-risk remedial students.

Of particular interest to Woolfolk (and others) is teacher efficacy during pre-service years and in the initial years of teaching. In their research involving 182 liberal arts majors enrolled in teacher preparation programs, Woolfolk & Hoy (1990) utilized a multiple-regression analysis to study teachers' sense of efficacy and classroom management beliefs. They found that GTE and PTE were appropriate dimensions of efficacy for pre-service teachers. They continued to identify two subareas within PTE: responsibility for student outcomes, and responsibility for negative outcomes. In this study, they found that teachers with high teaching efficacy were more humanistic in their student management. Also discovered was that GTE and PTE sometimes had opposite orientations when compared with the same dimension. For example, PTE correlated positively with bureaucratic orientation, and GTE correlated negatively. This study also brought to light that because GTE and PTE may correlate differently, it is important to account for the multiple permutations of these when studying potential relationships: high PTE/low GTE, low PTE/high GTE, high PTE/high GTE, as well as low PTE/low GTE. The prospective teachers who were the subjects of the study had spent no time in the classroom and were not equally represented by gender.

In Woolfolk-Hoy's 2000 study, "Changes in Teacher Efficacy During the Early Years of Teaching," she describes the beginning years of teaching as the most "malleable" in the teacher's career, and a teacher becomes less flexible as he or she becomes increasingly established. According to Woolfolk-Hoy, undergraduate pre-service teachers who possess a low sense of teacher efficacy tend to have a control orientation, view student motivation negatively, and rely on strict classroom regulations with extrinsic rewards and punishments. Student teachers who possess strong personal teaching efficacy were rated more positively by supervising teachers on lesson-presenting behavior, classroom management, and questioning behavior. Coursework and

pre-service practicum experiences impact levels of efficacy. General teaching efficacy tends to increase during college coursework and then decline during student teaching; this finding is consistent with research of Woolfolk and Hoy (1990). Student teachers' efficacy beliefs are impacted by the type of experience the students encounter. "Sink or swim" approaches are often met with poor outcomes and decreases in self-efficacy for the teachers. These teachers have management problems resulting in low efficacy and sometimes have difficulty responding as teachers and not as peers, which affects student efficacy.

The Woolfolk-Hoy 2000 study used multiple scales, including the Bandura form (1997), the OSU model (Tschannen-Moran & Hoy, 2001), and elements from the Gibson and Dembo form (1984). The study concluded the following:

- Efficacy generally rose during teacher training, but fell with actual teaching experience.

- On the Teachers' Sense of Efficacy Scale, an increase in the confidence factor correlated with a decrease in sick days taken.

- Teachers who encountered more sheltered support during their student teaching or first year of teaching showed positive changes in efficacy as measured by the Bandura and PTE scales.

The sample for this study was limited to 53 pre-service teachers at a single university. The two cohorts being compared were randomly assigned. In light of the small and not particularly representative sample, the study cannot easily be generalized. The Bandura self-efficacy items are said to correlate positively with GTE, PTE, and the OSU models, although Woolfolk-Hoy provides no numerical value to clarify the degree of correlation.

Gibson and Dembo (1984) found that teachers who have strong efficacy beliefs create mastery experiences for their students and show more persistence in failure situations, while

those who do not have strong efficacy beliefs are likely to undermine students' success and may blame the student and/or environmental factors for lack of success. The study has some shortcomings. The measure of GTE may not be consistent with Bandura's theory, given that Bandura speaks to the specificity of efficacy to a situation (1997). Since this study, the Gibson and Dembo (1984) assessment tool has come into question, which is not surprising, given that Gibson and Dembo (1984) themselves suggest that construct validation should continue to be investigated. Looking at observable classroom actions and activity as well as teacher decision making could have enhanced the study. The study also used a limited sample, so it may not be easily generalized. In fact, Bandura in his 1997 book refers to the Gibson and Dembo study as a microanalysis (p. 241).

Woolfolk and Hoy concluded that teachers with low efficacy have been found to prefer a cynical view toward student motivation, believe in strict regulation of the classroom, rely upon external inducements and sanctions in attempts to motivate students, and have classroom problems (Woolfolk & Hoy, 1990). The 1990 study highlights some of the challenges with measuring efficacy in that era, meaning that the validity is questionable. Additionally, the questions were mostly worded so that the responses that demonstrated high efficacy were negative responses requiring the subject to disagree. The PTE responses that demonstrated high efficacy were consistently affirmative responses requiring the subject to agree. This may have created bias, as well as confusion due to the need to disagree with negative statements to indicate a positive response.

Most importantly, teacher efficacy has been shown to positively correlate with student achievement (Anderson, Greene, and Loewen, 1988; Ashton & Webb, 1986, Vieluf et al., 2013). Ashton and Webb's study (1986) is a mixed-method study, which is logical given the timing of

the study with reference to the then recent conceptualization of teacher efficacy by Bandura. The sample was at-risk remedial students, and the relationship between teacher efficacy and student math performance was scant. The study by Anderson et al. (1988) shows significant correlation between PTE and academic performance on the Canadian Achievement Test for third graders, but not sixth graders. They attribute this to the belief (anecdotally shared by teachers) that by sixth grade, the teacher has less influence on whether a student performs than in younger grades. The scope, and thus generalizability, of the study is limited in that it included only two elementary grades. The participants of the study were 24 teachers who were strategically chosen due to their PTE scores, both low and high, allowing for easier comparisons. The sample selection process included surveying 77 teachers and choosing the 12 highest and lowest for personal efficacy, so long as their overall efficacy scores were above or below the mean, respectively.

Teachers' sense of efficacy predicted achievement on the Iowa Test of Basic Skills according to Moore and Esselman (1992). The study compared entire schools' levels of efficacy and student performance. The reason cited by the authors for using this comparison was related to the need to make individual teachers' responses anonymous due to the sensitive reactions to desegregation at the time. As a result, there exists potential for spurious influences. The sample size included 1,802 teachers who actually returned the instrument; however, the study does not indicate the number of teachers who received the survey. The study did find stability in the responses of the teachers, denoting less than 5% variance during a five-month period. This contradicts Bandura's assertion that self-efficacy is highly task-specific (1977). Due to the result-combining method, it could be argued that the collective efficacy of schools in Moore and

Esselman's (1992) study correlated with overall achievement, but that individual teacher self-efficacy's relationship to achievement was still in question.

According to Woolfolk, in an interview regarding her work with the concept of teacher self-efficacy, "Self-efficacy is the most useful self-schema for education because it relates to choices and actions that affect learning such as goal setting, persistence, resilience effort, and strategy" (Shaughnessy, 2004, p. 172). That being said, it would be naïve to believe that improving efficacy beliefs of teachers, in isolation, would make a difference.

An Australian study (Davies, 2004) found that teachers who had higher efficacy beliefs were more likely to hold higher order instructional objectives and outcomes. The students were also found to have higher success with higher order objectives. This work focused on individual teacher efficacy, rather than collective efficacy. This study was qualitative, based on interviews with teachers and students.

Teachers' sense of efficacy has also been linked positively to student motivation (Midgley, Feldlaufer & Eccles, 1989). In a longitudinal study of 1,329 students who experienced the transition from elementary to junior high school, it was discovered that students who moved from high-efficacy teachers in elementary school to low-efficacy teachers in junior high had the lowest self-expectations, even lower than those who had low-efficacy teachers at both levels. Teacher efficacy was more predictive for the perceptions of low-performing students than of high-performing students. However, the study focused on only a single grade-level transition, causing challenges with the ability to generalize these findings to other grades.

Klassen et al. (2011) conducted a thorough review of all peer-reviewed articles regarding teachers' sense of efficacy between 1998 and 2009. The purpose was to identify whether the nature of research had changed since Tschannen-Moran had questioned the entire construct in her

1998 publication. Of the 218 articles studied, only two during that time frame investigated teacher self-efficacy and actual student outcomes, and the relationships within those two articles were only modest. In short, the quantity and quality of published studies connecting teachers' sense of efficacy and student outcomes remains meager.

Several recent dissertations do show a link between efficacy levels and math achievement of students. Mills (2009) found a significant relationship between schools' collective efficacy scores and achievement of third graders on state exams in mathematics. This study includes all teachers within 15 schools, but it investigates collective efficacy, which shall be discussed later, rather than individual teacher efficacy.

Maguire (2011) found that teacher efficacy in student engagement and classroom management predicted differences in mathematical performance of students. This study included three high schools and utilized the Ohio State University Teachers' Sense of Efficacy Scale. This study did not investigate mathematics teaching efficacy specifically and may not be able to be generalized to an elementary school setting.

A 2006 study of teacher efficacy did not find a significant relationship between efficacy levels and student achievement, but it did find a relationship between efficacy levels using the Teachers' Sense of Efficacy Scale (Tschannen-Moran and Hoy, 2001) and principal evaluation ratings using the Danielson Evaluation Framework (Heneman et al., 2006). This finding points to the possibility that teacher self-efficacy has a relationship on the broad construct of teaching effectiveness, but also speaks to the difficulty in assessing a direct tie with student achievement, given the multitude of variables in a classroom.

Although some studies have shown that self-efficacy impacts achievement and classroom management, with over 30 years since Bandura first wrote about self-efficacy, the lack of studies

attempting to link teacher self-efficacy to achievement is surprising. Perhaps there exist so many confounding variables that a definitive link is not able to be found. An alternative is that self-efficacy levels vary enough, due to their situational nature, to make the correlation almost too challenging to find. Klassen et al. (2011) continue to suggest this as an area for future study, giving hope that the community of scholars may spend effort in this area.

Another area of research regarding teacher efficacy, which the Moore and Esselman (1992) study alluded to, involves the collective efficacy of a school. According to Bandura, the staffs of successful schools have a strong sense of efficacy to increase achievement in spite of difficult demographics or situations (Bandura, 1986, 1997). A widely cited study regarding collective teacher efficacy (Goddard, Hoy, & Hoy, 2000) indicates that schools with strong collective efficacy are less likely to give up on students, and that schools with low collective efficacy are less likely to accept responsibility for low achievement, blaming student risk factors. The study involved a limited sample from elementary schools in the Midwestern United States. In the study, the researchers designed an instrument to measure collective efficacy. The measure was found to have strong reliability and reasonable validity. Schools with higher collective efficacy correlated with those whose students scored more favorably in mathematics and reading (Goddard, Hoy, & Hoy, 2000). That being said, given that it is difficult to control this type of study, other influences may be at play.

Influences on Teachers' Sense of Efficacy

Bandura (1994) describes four sources of self-efficacy that can easily be expanded to the construct of teacher self-efficacy: mastery experiences, vicarious experiences, social persuasion, and reduction of people's physiological stress reactions. He continues to discuss lifespan influences on efficacy, with attention to life stages, which is similar to the teachings of Piaget. He

indicates that self-efficacy is initiated as an infant by realizing that one can improve one's situation by taking actions to help oneself and receiving personal reward. Family provides reinforcement of self-efficacy by providing play activities, helping children develop linguistic competence, and teaching social competence. As students grow older, feedback and information from peers impacts self-efficacy. School is a primary source of social confidence and cognitive competence. Through adolescence, experiences, experimentation, peer interactions, risk-taking, and increasing exposure to adult-like decisions can affect self-efficacy. For adults, marital relationships, career experiences, and parenthood provide efficacy impacts. With advancing age, self-reflection affects one's self-efficacy beliefs (Bandura, 1986).

Gist and Mitchell (1992) report determinants and malleability of self-efficacy, primarily related to the field of business, rather than education. The report provides many suggestions for use of Bandura's theory; however, it is not a study. The contents, however, do appear to be generally consistent with social cognitive theory. They suggest that when a person's capability is high but the individual incorrectly self-assesses, believing that his or her capability is low, positive information and reassurance can lead to great increases in self-efficacy. They also found that where low self-efficacy is an accurate reflection of likely performance, verbal persuasion is likely to decrease self-efficacy if failure occurs. This is likely a problem with the construct of self-esteem. Educators often fall into the trap of encouraging students or one another, regardless of whether the student or teacher has the requisite skills; this can actually decrease self-efficacy, thereby resulting in a downward spiral. Gist and Mitchell's (1992) third finding is that the ability to perform a task is sometimes more important than efficacy beliefs. The fourth finding is that attribution processing and feedback affect efficacy, because self-efficacy mediates the translation of knowledge and abilities into skilled performance.

Attribution is the individual's perception of what actually causes results (e.g. hard work vs. luck). Feedback regarding task performance can assist with helping one to increase or decrease confidence in the ability to improve results. Gist and Mitchell suggest that to help an individual improve performance, a trainer can (a) provide more information regarding the task attributes, complexity, task environment (using modeling), and how to control these variables, (b) provide training to improve one's abilities, and (c) help one to better understand the psychological or effort expenditure requirements to complete the task. Although these strategies are relevant to employees in the business world, professional developers may be well served to utilize the strategies when training, mentoring, and supporting teachers (Gist & Mitchell, 1992). Developmental influences that exist for other professionals impact teachers as well.

There are additional impacts specific to the self-efficacy of teachers: one's own mastery experiences, vicarious experiences, and social persuasion (Bandura, 1977). Early in the career, positive experiences help increase the future outcome expectations and thus increase the perceived efficacy of the teacher. However, the converse is also true; failure in the early years can lead to feelings of helplessness and to lower efficacy perception. After more experience, the teacher may be less impacted by failure and may be more willing to take risks if her or his overall efficacy is strong. Besides personal experiences, the experiences of others, such as colleagues, that the teacher is aware of - vicarious experiences - affect efficacy levels. The closer the observer identifies with the model, the more likely the impact of this modeling. Social persuasion, the third specific impact to teacher efficacy, may constitute informal conversations among teachers or supervisory feedback. To the degree the credibility of the persuader is strong, the perception of efficacy will be affected (Woolfolk-Hoy, 2000).

Ross's 1995 study of ways to enhance teacher efficacy cites concern regarding the threats to validity and reliability of many inferential studies regarding teacher-efficacy, including those of Woolfolk and Hoy. He cites concern regarding spurious impacts and the difficulty of properly controlling studies of in-service training and concludes that whether in-service training affects teacher self-efficacy is inconclusive. Ross (1995) suggests that there is more evidence supporting the benefits of blended training in which skills, reflection, collaboration, and belief structures are involved. Models include peer coaching, vertical teaming, and/or mutual training coupled with use of student performance data. There are, however, risks to the blended approach. Data might indicate a lack of results, which could reduce teachers' efficacy. Collaborative professional development, if it lacks direction, may not improve results for students or efficacy. Ross (1995) also cites negative efficacy effects caused by externally imposed changes, including state-initiated school reform initiatives and merit pay (or career ladder) systems.

Of particular interest to policymakers and practitioners, however, is whether there is a relationship between content area knowledge and efficacy. Another question is whether there is a relationship between professional development and teacher self-efficacy. This research shall explore each with an emphasis on mathematics.

Mathematics Training and Teachers' Sense of Efficacy

Recent research regarding mathematics content knowledge, typically measured by number of college credits in mathematics, and self-efficacy levels of teachers have varied in their conclusions. One study of elementary school teachers found that the teachers with the lowest efficacy were the teachers whose highest mathematics class in college was a remedial class (Carson, 2012). The same study also found that the teachers with advanced mathematics coursework in college showed strong efficacy in the teaching of geometry. A 2012 study of high

school teachers found no significant difference between number of college courses in mathematics and levels of efficacy (Jones). A 2011 study of pre-service elementary school teachers (McCoy) found significant positive correlations between personal teaching efficacy levels and possession of specialized mathematical knowledge, but it did not find significant correlations between mathematics teaching efficacy levels and specialized mathematical knowledge. In a qualitative study (Phelps, 2009) teachers also attributed their own mathematical knowledge, among other factors, to their sense of efficacy in teaching of mathematics. A study comparing alternatively credentialed teachers in Florida (Elliott, Isaacs, & Chugani, 2010) demonstrated a significant positive difference in self-efficacy levels of traditionally certified teachers as compared with alternatively certified teachers. A positive significant relationship was found between participation in mathematics and science courses and self-efficacy for teaching these subjects for middle school teachers (Swackhamer, Koellner, Basile, & Kimbrough, 2009). A longitudinal analysis found that pre-service content knowledge delivered through the means of a university pre-service methods course significantly and positively related with gains in personal teaching efficacy, but not in outcome expectancy (Newton, Leonard, Evans, & Eastburn, 2012).

Multiple studies have demonstrated that delivery of professional development for elementary teachers can influence teachers' sense of efficacy in mathematics instruction. Zambo and Zambo (2008) found that teachers' efficacy beliefs showed statistically significant increases in personal mathematics confidence after teachers participated in an intensive training in both problem solving and teaching methods. Another qualitative study found that teacher self-efficacy improved through participation in a teacher support program focused on a support network, idea sharing sessions where strategies where shared, and observation of successful veteran teachers.

In this study, elementary teachers who participated in a multi-week embedded reform-based training consistent with methods recommended by the National Council of Teachers of Mathematics reported increases in their sense of efficacy and habits of self-reflection, and their supervisor evaluations improved (Bruce & Ross, 2008). Three studies have found that an embedded, ongoing, professional development program in mathematics that includes actual mathematical practice, dialogue about strategies for teaching, teacher observations, and coaching increased levels of efficacy for teaching mathematics (Aerni, 2008; Althauser, 2010; Tschannen-Moran, & McMaster, 2009). Mathematics in-service professional development as an approach to impact efficacy appears to have promise if implemented in an embedded manner.

Isiksal (2005) identified a relationship between teachers' pre-service performance during the yearlong mathematics education preparatory programs at the university and later levels of self-efficacy for teaching mathematics. This finding was reinforced when Isiksal & Cakiroglu (2005) identified, in a separate study, a positive relationship between mathematics teaching efficacy levels and academic performance in university coursework for pre-service teachers. In a third study, Isiksal (2010) identified an inverse relationship between mathematics teaching efficacy and math anxiety for new teachers.

During the last decade, scholars have identified a construct entitled *mathematical knowledge for teaching* (Ball, Hill, & Bass, 2005; Hill, 2010; Hill, Rowan, & Ball, 2005; Hill et al., 2008). In contemplating how to best prepare mathematics teachers, scholars have found that mathematical content knowledge is not sufficient to prepare the teacher. A higher yield attribute of teachers identified in the literature is mathematics knowledge for teaching (MKT). The two primary ingredients of MKT are mathematical knowledge common to individuals working in diverse professions, and subject matter knowledge that supports mathematics teaching. Examples

of subject-matter knowledge that supports teaching include knowing how and why procedures work, how to define mathematical terms for appropriate grade levels, and the types of errors students are likely to make. Teachers who possess MKT are likely to deliver a higher mathematical quality of instruction (MQI). MQI includes quantity of mathematical errors, explanations and justifications, and how one represents mathematics symbolically. This body of research not only suggests promising attributes for teacher preparation and professional development, but also likely has implications for self-efficacy levels of teachers, because mathematics-specific pedagogical knowledge predicts personal teaching efficacy (Newton et al., 2012).

Value-added Modeling

Teacher evaluation in the state of Florida was changed legislatively in 2011 (Florida Statutes Annotated § 1012.34(3)(a)(1), 2011), to require enhanced teacher evaluation systems and a value-added element. Value-added models identify growth in aggregate test scores from year to year for the students assigned to an individual teacher. The teacher's aggregate score is compared to the growth of student scores state wide, called expected growth. If the growth scores of the teacher's class exceed the state-expected growth, then the teacher effect is denoted as positive. Conversely, if the scores of the teacher's class are more meager than the state-expected growth, then the teacher effect is denoted as negative. Factors that are considered within the formula are class size, attendance, and students' previous scores. An excluded element is whether the students are subject to poverty, a consistent predictor of school success (Florida Department of Education, 2012).

According to § 1012.34, starting in 2014, at least half of a teacher's evaluation must be based on his or her value-added scores. The other portion is based on teacher observation scores

yielded from the Marzano system (2011). The overall evaluation scores are high stakes, allowing

for contract non-renewal for a series of Needs Improvement or Unsatisfactory evaluations.

Additionally, teacher performance pay in Florida is tied to evaluation results. The Florida

Education Association and affiliated teachers have raised a number of legal challenges to the

methodology, implementation, and value-added computation. Such challenges have thus far

yielded little change in the system. The Florida model is now being analyzed within legal

publications, indicating that the controversy has not reached a conclusion (DeMitchell,

DeMitchell, & Gagnon, 2012).

 A number of educational and legal scholars advise caution when policymakers consider

value-added modeling for high-stakes decision making (Amrein-Beardsley, 2008; Braun, 2005,

2013; DeMitchell et al., 2012; Dively, 2012; Hanushek & Rivkin, 2010; Hill & Grossman, 2013;

Hill, Kapitula, & Umland, 2011; McCaffrey et al., 2003; Newton, Darling-Hammond, Haertel, &

Thomas, 2010; Rothstein et al., 2010). Such studies criticize the model based on equity

arguments, lack of random student assignment, oversimplification, and lack of consideration of

sufficient variables. A typical example of such arguments is the following:

> …there is broad agreement among statisticians, psychometricians, and
>
> economists that student test scores alone are not sufficiently reliable and valid
>
> indicators of teacher effectiveness to be used in high-stakes personnel
>
> decisions, even when the most sophisticated statistical application, such as
>
> value-added modeling is employed. For a variety of reasons, analyses of VAM
>
> results have led researchers to doubt whether the methodology can accurately
>
> identify more and less effective teachers (Rothstein et al., 2010, p. 7).

On April 8, 2014, the American Statistical Association (ASA) issued a position statement regarding value-added measures in education (American Statistical Association, 2014). The statement offers many cautions. The ASA statement warns against using value-added modeling for high-stakes decisions. It also reminds scholars that value-added measures identify correlation, but not causation. They caution that one should not necessarily attribute measured results (such as test scores) to teacher effectiveness. The statement indicates that value-added metrics should be used for organizational improvement, but teachers should not be ranked based on the metric. The statement also offers caution that use of value-added models may have unintended consequences that may be harmful to teachers, students, or the education profession.

Teacher evaluation systems that utilize value-added metrics are a relatively new phenomenon. One study has found a weak but positive correlation between value-added scores for teachers and classroom observations by principals (Harris, Ingle, & Rutledge, 2014). A recent mixed-methods conference paper sponsored by the Gates Foundation (Goldring et al., 2014) suggests that improved classroom observation systems have more impact on school and teacher improvement momentum than value-added scores.

Despite the lack of clear evidence linking value-added evaluations to student achievement, the number of states electing to implement policy using value-added growth data as the preponderant criterion in teacher evaluations has increased from 11 in 2012 to 20 in 2013 (Doherty & Jacobs, 2013). With this rapid increase, an initial study linking self-efficacy and value-added scores was published as a dissertation in 2014. This grounded theory study (German, 2014) identified four themes connecting value-added ratings and efficacy: self-beliefs, perceived inaccuracies of value-added methods, promising practices, and a desire for privacy. The rapid growth in utilization

of value-added models and initiation of qualitative research on the topic make value-added approaches ripe territory for quantitative research (German, 2014).

Marzano Evaluation Model

Robert Marzano's "Causal Evaluation System" (Marzano, Frontier, & Livingston, 2011) lists 41 teaching strategies that tend to correlate with student achievement. The identification of these strategies was achieved through Marzano's meta-analyses, which are clarified in his book entitled *The Art and Science of Teaching* (2007). Subsequent to publication of this book, Marzano's team created a teacher evaluation system that adds non-observational criteria to create a comprehensive view of the teacher, which includes a total of 60 elements (Marzano et al., 2011; Marzano & Toth, 2013). The elements are denoted in what he calls a *Learning Map* (Appendix C). The trained principal observer is to identify dominant elements from the evaluation system. Dominant elements are those that are germane to the particular lesson that is being observed, because it would be impossible to observe all 41instructional elements simultaneously. Additionally, many elements may not be applicable to the lesson being taught while the observer is there, either due to the limited time frame the observer is in the teacher's room, or because the lesson does not lend itself to those strategies.

After spending time in the classroom, the evaluator scores each dominant element using a scale. The scale is nominal in nature. The scoring descriptors are as follows: innovating, applying, developing, beginning, and not-using (Marzano & Toth, 2013). When the ratings over a series of observations are combined with the non-instructional elements, an aggregate annual score is identified. Each school district in the state has a slightly different method of determining the aggregate score due to local

collective bargaining, but the commonality is that a predominant collection of applying

and innovating scores, when combined with few not-using, beginning, and developing

scores, will yield an "Effective" or "Highly Effective" total evaluation. Conversely, a

predominant collection of developing and not-using scores, with few innovating,

applying, or developing scores, will yield a Needs Improvement or Unsatisfactory final

evaluation rating.

Although the Marzano system is relatively new, one recent article points to

needed enhancements. Hill and Grossman, in their review of the current evaluation

systems (2013), note a need for the systems to be used to assist with professional

practice rather than as a summative review. They also note that the systems lack subject

specificity, and sometimes completely miss the essence of particular content areas. They

suggest the need for content-area specialists to conduct the evaluations—an element

missing in many schools due to the scope of principal supervisory duties. Also

suggested is that information must be accurate, timely, and useful for teachers. No

literature, to date, appears to link teacher self-efficacy and the Marzano model.

Conclusion

As has been the case for the last century, mathematics instruction is a prime area of

concern and focus for educators and policymakers (Abbott et al., 2010; Mewborn, 2013; Nisbet

& Warren, 2013; Wilson, 2013). Wilson (2013) claims, "As states raise the bar on math, students

are learning more complex subjects, earlier on in their educational lives. This trend has presented

particular problems for some elementary school teachers, whose pedagogy has typically centered

on literacy" (p. 2). Continued investigation of the relationship between mathematics teaching,

elementary teachers' sense of efficacy, and teacher training appears warranted given the limited research and mixed results of existing studies on the topic.

Most studies connecting professional development methods and teacher self-efficacy are qualitative in nature, so a quantitative study will be helpful to increase the body of knowledge. The limited number of studies relating teachers' sense of efficacy and student performance, despite calls for further research in this area, makes this area ripe for research (Klassen et al., 2010).

Chapter III

Design and Methodology

Introduction

This study included several correlational approaches. The independent variables were two different measures of teachers' sense of efficacy: the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Hoy, 2001), and the more content-specific Mathematics Teaching Efficacy Beliefs Instrument (Enochs et al., 2000). Both are surveys that are administered to teachers to determine their self-efficacy beliefs. The dependent variables were value-added scores using the Florida value-added metric (Florida Department of Education, 2012), teacher evaluation scores using the Marzano evaluation model (Marzano, 2013), and overall evaluation scores identified by the school district.

Subjects were fourth- and fifth-grade teachers from a school district in Florida that represents state demographics. The school district included 110 fourth- and fifth-grade teachers, and 32 teachers were included in the sample. This number is sufficient for correlational methods (Bartz, 1976; Metzler & Charles, 2011).

To answer the first research question a point-biserial correlation was used because the predictor variables were derived from Likert scales (Salkind, 2013). For the second question a single regression was used, because the predictor and resultant variables were scale scores. For the third research question, there were five predictors and the resultant variable was ordinal, so a multiple regression was used (Field, 2013).

Research Design

This quantitative study analyzed fourth- and fifth-grade teachers' sense of efficacy for mathematics instruction. Two elements investigated in the study were how teacher self-efficacy

(TSE) correlates with teacher evaluation scores using the Marzano model, and how TSE

correlates with teachers' students' value-added mathematics scores derived by the State of

Florida. A third element was regression relationships between the evaluation ratings for

mathematics instruction with (a) general teacher TSE, (b) mathematics instruction TSE, (c) years

of teaching experience, (d) number of mathematics courses taken by each teacher, and (e) recent

hours of mathematics professional development for each teacher (see Figure 2).

Figure 2

Study Design
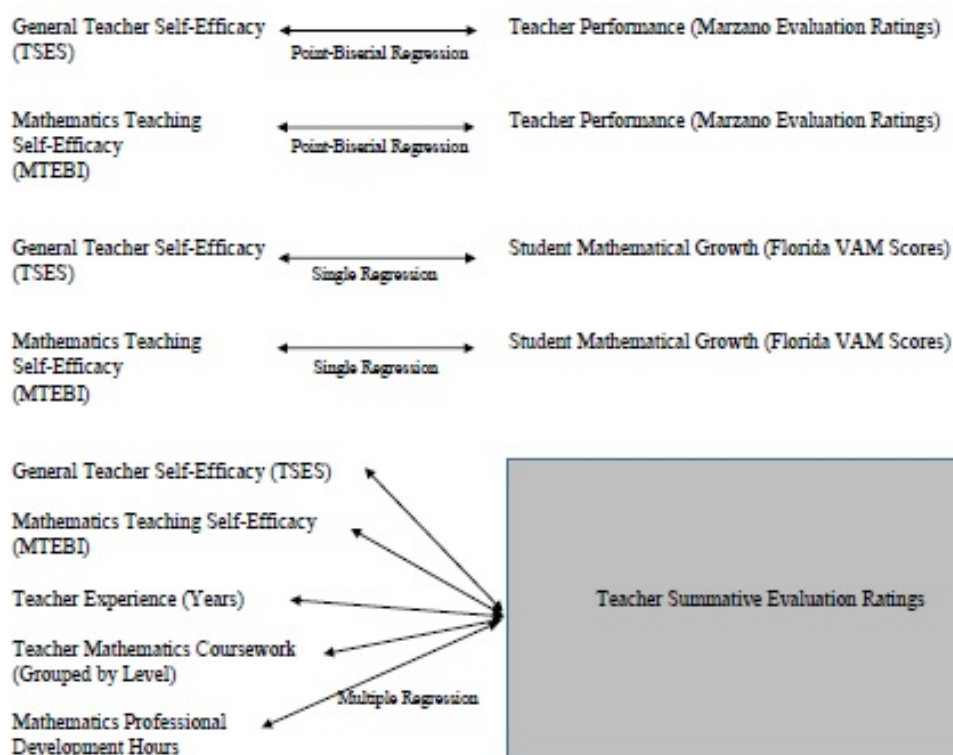
Note. Study design for the dissertation describes the independent variables on the left and the dependent variables on the right, with the statistical regression method denoted near the multidirectional arrows in the center, corresponding to each research question.

The research questions explored in this study included:

1. What is the relationship between teachers' levels of self-efficacy and teacher observation scores derived from the Marzano evaluation system?

2. What is the relationship between teachers' sense of self-efficacy and Florida State value-added scores in mathematics?

3. What proportion of influence do the following have on a teacher's evaluation rating: teacher sense of efficacy, as measured by the Tschannen-Hoy instrument; mathematics teaching efficacy, as measured by the MTEBI; years of teaching experience; mathematics professional development level; and mathematics coursework attained?

Participants

The sample of teachers chosen for this study involved 32 fourth- and fifth-grade teachers from a school district in the state of Florida. All teachers were charged with teaching mathematics, although most were responsible for teaching other subjects as well. Participant age ranged from 21 to 62 years ($\mu = 42.06$ years). Years of teaching experience ranged from 1 to 34, with a mean of 15.25 years. Teacher age and years of experience each followed a normal pattern. In this study, 12.5% of the subjects identified as male, 87.5% identified as female, and one did not identify his or her gender. The self-identified ethnic composition of the subjects was 93.8% Caucasian, 3.1% African-American, and 3.1% Hispanic/Latino. The relationships between demographic data for the study group, district group, and state group are denoted in Appendix A. This analysis utilized the self-reported demographics of the study group and 2013-14 state faculty demographics (Florida Department of Education, 2014b). Whether the sample school

district demographics and state demographics are similar is irrelevant because self-efficacy has been shown to not vary significantly based on teacher race (Tschannen-Moran & Hoy, 2007).

The school system studied had a student population of approximately 17,000 students. In the county in which this study took place, 57.17% of the school-age children were white (non-Hispanic or Latino), 21.24% of students were Hispanic or Latino, and 16.45% of students were African-American (US Census Bureau, 2010). In 2010, 65% of students qualified for free or reduced lunch in this county (Florida Department of Education, 2011). The school district comprised an urban area, which included a business center and higher density housing; suburban subdivisions, varying in housing style and income level; and rural areas, which included farm workers and estate owners. The school district also included an area along the Atlantic Ocean with large homes and relatively low student density (due to retirees and seasonal ownership). Of the 13 schools studied, one was located next to the ocean, three were in the urban area, eight were in the suburban area, and one was in the rural region. Consent to conduct the study was obtained from the Superintendent of Schools (see Appendix D).

Data Collection

The study included three research questions. This approach allowed for a holistic analysis of elementary teachers' sense of efficacy, including self-efficacy and mathematics instruction, self-efficacy and student achievement growth in mathematics, and relationships between self-efficacy, coursework of teachers, teacher experience, and evaluation scores. Each question suggested a different design approach.

The study included 32 fourth- and fifth-grade teachers from 13 schools. A sample size smaller than 30 is not likely to accurately reflect the trait distributions that exist in the population from which it was drawn (Mertler & Charles, 2011). Minimum sample sizes are dependent upon

the type of research investigation undertaken. In the case of correlation studies, statisticians generally agree upon a threshold of 30, making the study's sample size of 32 sufficient. More important than the size of a sample is the accuracy of the sample selected for a given investigation (Bartz, 1976).

Informed consent was obtained from the subjects (see Appendix E). Within the consent process, teachers were informed that their information would be coded and their evaluation scores would be compiled, but anonymity would be maintained through a coding system as described by Creswell (2012). The efficacy levels for the study were gathered using an online survey, which included demographic elements and questions from both the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Hoy, 2001) and the Mathematics Teacher Efficacy Beliefs Instrument (Enochs et al. 2000). Qualtrics (2014) survey software was used to administer the questions. The survey results were gathered over a three-week period during the fall of 2014. Reminders to the subjects were sent out at the start of the second and third weeks of the study to increase participation (see Appendix F). The survey was sent to all teachers, but data were gathered from only 32 participants by the end of the three-week survey period.

Demographic information was gathered from the teachers through questions on the survey (see Appendix G). Questions solicited the subjects' gender, race, years of teaching experience, age, and level of mathematics schooling. The mathematics schooling data categories for the survey were clustered based on level of schooling. Because the names of college classes may be varied depending on the university, the survey asked participants to identify any college classes they had taken, and ordinal clustering was determined by the researcher. Although not the primary focus of this study, the information was gathered to see whether any salient trends existed for these factors.

For each of the three research questions, teacher self-efficacy was measured. Levels of general teacher self-efficacy were measured using the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Hoy, 2001), and the Mathematics Teaching Efficacy Beliefs Instrument (Enochs et al., 2000). The Teachers' Sense of Efficacy Scale is a commonly used instrument to measure teacher self-efficacy, and multiple analyses have confirmed its construct validity (Heneman et al., 2006; Klassen et al., 2011). This tool, however, is not specific to the content area of mathematics. The Mathematics Teaching Efficacy Beliefs Instrument (Enochs et al., 2000), was also used due to its content specificity and validity (Cetinkaya & Erbas, 2011; Enochs et al., 2000). Recently, the MTEBI's construct validity has begun to be questioned (Kieftenbeld, Natesan, & Eddy, 2010). Although this matter shall be discussed in greater detail later in this study as a limitation, it is important to note that the critique has been made within the literature. A new tool is suggested for development (Ward & Johnston, 2014). However, the MTEBI is currently the only teacher self-efficacy measurement tool specific to mathematics.

Each of the tools used in this study uses a Likert scale, yielding an ordinal measure for each question (Creswell, 2012). The Likert data are then, for each tool, converted into a scale score. This approach is questionable, because changing ordinal data into a scale score is not standard statistical procedure. However, the originators of the instruments provide instructions for making such a conversion, and the prevalent use of these tools in this way has yielded what are agreed upon as valid and reliable results (Bakar, Mohamed, & Zakaria, 2012; Domsch, 2009; Esterly, 2003, Fives & Buehl, 2010; Head, 2012; Heneman et al., 2006; Jones, 2012; Tschannen-Moran & Hoy, 2001). Additionally, treating ordinal data as a scale score becomes statistically appropriate when there are more than five degrees in the Likert scale (Johnson & Creech, 1983). Because of this mild dilemma, the data were treated as scale scores for the purposes of the

different regression methods in the study, but raw data were reported for the Likert data (see Appendix H), and any anomalies are noted in Chapter IV.

The Teachers' Sense of Efficacy Scale is authorized for open use; a letter documenting such is included in Appendix I. Permission was procured for use of the Mathematics Teaching Efficacy Beliefs Instrument from its authors (see Appendix J). The typical loading factors are noted in Table 1.

Table 1

Data Collection Tools

Instrument Name	Abbreviation	Author	Typical Factor Loading
Teachers' Sense of Efficacy Scale	TSES	Tschannen-Moran & Hoy, 2001	Efficacy for Student Engagement Efficacy for Instructional Practices Efficacy for Classroom Management
Mathematics Teaching Efficacy Beliefs Instrument	MTEBI	Enochs, Smith, & Huinker, 2000	Personal Mathematics Teaching Efficacy Mathematics Teaching Outcome Expectancy

A number of previous studies document the validity of both teacher self-efficacy measurement tools. The Cronbach's alpha levels from several studies are noted in Table 2. Alpha levels between .70 and .90 are generally considered to be acceptable indicators of internal validity (Tavakol & Dennick, 2011). Given that multiple studies verify alpha levels in the acceptable range, the research tools can be considered valid for this study. A confirmatory factor analysis was conducted to verify the validity of the measures and the internal factors, and the results are discussed in Chapter IV.

For each teacher, three other pieces of data were gathered from the school district and coded to maintain anonymity while allowing the information to be linked to the survey results. The three data elements were each teacher's Marzano observation score, each teacher's Florida value-added measure (VAM) score in mathematics, and each teacher's overall mathematics evaluation score. Each teacher's Marzano score and VAM score were used for research questions 1 and 2.

Table 2

Typical Validity Measures for Teachers' Sense of Efficacy Scale and Mathematics Teaching Efficacy Beliefs Instrument

Tool	Researchers and Study	Overall Validity	Elements' Factor Validities			*n*
			Instruction	Management	Engagement	
TSES	Tschannen-Moran & Hoy, 2001, p. 800	0.90	0.86	0.86	0.81	366
TSES	Heneman et al., 2006, p. 18	0.89	0.77	0.90	0.84	1,075
TSES	Klassen et al., 2009, p. 74	0.87	0.76	0.88	0.82	137
TSES	Head, 2012, p. 69	0.93	0.83	0.85	0.83	26
TSES	Tschannen-Moran & Hoy, 2007, p. 15	0.93	0.87	0.88	0.84	255

Tool	Researchers and Study	Overall Validity	Elements' Factor Validities		*n*
			Personal Math TE	Math OE	
MTEBI	Enochs et al., 2000, p. 196	0.92	0.88	0.77	324
MTEBI	Farmani & Khamesan, 2011, p. 11	0.73	0.80	0.66	140
MTEBI	Althauser, 2010, p. 71	0.90	0.90	0.74	35
MTEBI	Head, 2012, p. 101	0.77	0.74	0.86	263
MTEBI	Kieftenbeld et al., 2011, p. 44	0.85	0.89	0.81	217

Notes: Validity scores are Cronbach's alpha measures. Althauser (2010): Scores are preservice scores from the study; Outcome expectancy is noted as "general efficacy" in the study but used the same questions as OE for the MTEBI instrument; n-value low for valid confirmatory analysis. MTEBI = Mathematics Teaching Efficacy Beliefs Instrument; TSES = Teacher Sense of Efficacy Scale

The final scores included within the study, used for question 3, were evaluation scores, which included a blend of 60% evaluation ratings from the principal evaluator using the Marzano evaluation model and 40% value-added scores in mathematics (Florida Department of Education, 2011). The scores were used to determine an overall rating for each teacher: Highly Effective, Effective, Needs Improvement, or Unsatisfactory. This variable, used only for the third research question, was an ordinal value. The dependent variable, therefore, was a composite measure, combining both teaching performance and student achievement growth (Copa, 2012).

A school district employee served as a surrogate to administer the survey instruments. The surrogate matched the data with the evaluation results in such a manner that the researcher did not have any personally identifiable information. This method was proposed because of the sensitivity surrounding teacher evaluation data in a high-stakes environment. The method of using a surrogate addresses the ethical consideration of privacy (Drew, Hardman, & Hosp, 2008). Conversely, using a surrogate would have masked the identity of the researcher, and one may question whether the approach is consistent with the ethical obligation to be open and honest (Drew, Hardman, & Hosp, 2008). When one weighs the ethical consideration of privacy against openness, privacy is the more impactful risk to the subjects. Additionally, validity was enhanced by using a surrogate because subjects, not knowing who the researcher was, were not likely to answer questions based on what they thought the researcher wanted them to say; thus one potential element of bias was removed.

Analytic Methods

Three different methods were used to analyze the data. Data analysis was conducted using SPSS. For research question 1, the independent variables were each teacher's level of efficacy as measured by two Likert-scale tools, which were converted into scale scores. The

dependent Marzano score variable, however, was nominal. When data are continuous in nature, and represented via the nominal and interval scales of measurement, the point-biserial correlation coefficient allows the researcher to best determine the relationship between the variables under investigation (Salkind, 2013).

For research question 2, the independent variables remained each teacher's two self-efficacy measurement tools, both converted to scale scores. The dependent variable for this question was the value-added score, which was an interval variable. When the researcher wishes to establish the mathematical relationship between two scale or interval variables, the statistic of choice is a single regression (Salkind, 2013).

For the third research question—the relationship between multiple factors and teacher evaluation scores—a multiple regression model was used. When the researcher's interest in an investigation is to determine the predictive robustness of independent variables, multiple regression is generally the statistic of choice (Field, 2013; Laerd Statistics, 2014).

Research question 1: What is the relationship between teachers' levels of self-efficacy and teacher observation scores derived from the Marzano evaluation system? For this portion of the study, the independent variables were the teachers' sense of efficacy, measured by the TSES (Woolfolk & Hoy, 2001) and measured again by the MTEBI (Enochs et al., 2001). Evaluation scores were obtained using the Marzano model (Marzano, 2013) as the dependent variable. The overall scores from 12 evaluation experiences (some formal, some informal, and some drop-in) were combined to create numerical composite performance scores for the each teacher. This question was addressed by using a point-biserial regression.

Research question 2: What is the relationship between teachers' sense of self-efficacy and their Florida State value-added scores in mathematics? For this portion of the study, the

independent variables were the teachers' sense of efficacy, measured by the TSES (Woolfolk & Hoy, 2001) and measured again by the MTEBI (Enochs et al., 2001). The dependent variable was each teacher's VAM scores (Copa, 2012, Braun, 2005). This question was addressed by using a single regression. The value-added scores for each teacher ranged from -4.0 to 4.0, with a score of 0 being the expected mathematical growth for the teacher's students, and the measures, either negative or positive, being z scores denoting how many standard deviations from the expected value the teacher's students' scores fell. In other words, a negative score meant less growth than expected, and a positive score meant more growth than expected in a given year. This question was addressed using a single regression.

Research question 3: What proportion of influence do the following have on a teacher's evaluation rating: teacher sense of efficacy, as measured by the Tschannen-Moran & Hoy instrument; mathematics teaching efficacy, as measured by the MTEBI; years of teaching experience; mathematics instructional professional development; and mathematics coursework attained? This portion of the study utilized a multiple regression analysis. The five independent variables tested were the TSES, the MTEBI, the teachers' years of experience, the teacher's mathematics professional development, and the teachers' levels of mathematics training in college. The dependent variable was each subject's summative evaluation score. This method allowed for identification of the variance of each predictor as well as the combined effect of the predictors.

Ethical Considerations

Throughout the design, implementation, and reporting phases of the investigation, ethics were addressed. Ethical elements included a) disclosure and consent, b) voluntariness, c) ethical

use of incentives, d) review of risks to subjects, e) costs and benefit of deception, f) institutional review., g) site permission, h) consent from other researchers and i) reporting of data.

Disclosure and consent. The study gleaned information from a survey as well as teachers' evaluation data. A surrogate researcher mailed the electronic survey request to all fourth- and fifth-grade teachers in the school district. All subjects were adult teachers and were capable of giving consent. The survey was described to the potential subjects in the electronic introduction, and subjects were invited to click on the survey link to participate.

The first page of the electronic survey was the Informed Consent Form (see Appendix E). The Informed Consent Form explains the purpose of the study, that a survey would be administered, that participants' evaluation results would be obtained by the researcher, the use of coding to keep identities of subjects anonymous, and that the subjects had no obligation to complete the survey if they decided to stop prior to completion. Many subjects appeared to understand that they could stop early; 46 subjects began the survey, but only 32 actually completed the instrument. By continuing, potential subjects provided consent to take the survey as well as to have their evaluation data released and linked to their survey through coding. The surrogate researcher sent two reminder notices to potential subjects to encourage participation (Appendix F).

Voluntariness. The Informed Consent Form was clear that participation was voluntary and there would be no repercussions for non-participation. The research surrogate served as a liaison between the researcher and the subjects. The liaison did not supervise any of the potential subjects, so there was no implicit requirement to participate.

Ethical use of incentives. As an incentive, potential subjects were informed that a drawing would be held for a prize to thank them participating. Two participants were selected by

the surrogate researcher to each receive a $50 gift card to a local supermarket. The researcher successfully completed the Protecting Human Research Participants training provided by the National Institutes for Health (NIH) (see Appendix K). The three primary considerations to consider with human research, consistent with the NIH training are beneficence, respect for persons, and justice. The beneficence test weighs the potential risk of participation with the benefit of the incentive (Grant & Sugarman, 2004; Singer and Bossarte, 2006). In the case of this study, the participatory risk was low, especially given the coding of data and anonymous reporting of results. Evaluation results are personal, and their release could have negative professional consequences. However, in Florida, evaluation results become a public record within a year of release to the teacher. The risk of a coded release to the researcher, although posing some potential privacy risk to the subjects, likely posed less risk than a legal release of the information to a local newspaper would pose. The incentive initiated by the researcher was participation in a drawing for two prizes of diminutive value, $50 worth of groceries. The low level of risk, when combined with the minimal value of the incentive, met the beneficence test. Respect, consistent with the Belmont Report (U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979) primarily involves the potential subjects' autonomy, which was addressed by administering informed consent and the absence of a vulnerable sampling pool. The issue of justice was addressed through the non-discriminatory administration of the survey. The study did offer an incentive, but did not meet the operational standard of being coercive or even the lesser standard of being manipulative (Grant & Sugarman, 2004; Singer and Bossarte, 2006).

Review of risks to subjects. The primary risks to subjects was release of personally identifiable survey or performance evaluation information. Use of a surrogate school district

employee to administer the survey, link the evaluation data to the surveys using a coding system, and provide the data to the researcher, devoid of any personally identifiable information, mitigated the risk of loss of privacy (Creswell, 2012; Drew, Hardman, & Hosp, 2008). At no point did the researcher know the personal identity of any subject. Additionally, data was reported in such a way as to protect the identity of individual subjects.

Costs and benefit of deception. When considering research ethics, one must consider the costs and benefits of deception (Drew, Hardman, & Hosp, 2008). The potential subjects for the study did not know the identity of the researcher. This omission decision was due to several factors, namely, anonymity of subjects and reducing any risk of perceived coercion. By identifying a school district surrogate to administer the survey and using coding to gather the evaluation results, the researcher provided an additional layer of anonymity for the subjects. The information was released to the researcher as a fully anonymous data file. The researcher holds an administrative position in a Florida school district that is responsible for administration of hiring and human resources oversight. If the researcher's identity had been known to the potential subjects, they may have perceived a level of coercion to participate or perceived that participation would somehow have a nexus to employment decisions. Through use of a surrogate who helped keep distance between the subjects and the researcher, the researcher avoided any perceived risk of coercion or conflict. The benefits afforded by using a surrogate researcher outweighed the costs.

Institutional review. Consistent with NIH requirements, Northwest Nazarene University administers an institutional review board, called the Human Research Review Committee. The committee reviewed the study and provided its approval (Protocol Number 2514931).

Site permission. The superintendent of the school district where the study took place was made aware of the scope of the study and plan for administration and provided written permission for the study to be conducted (see Appendix D).

Consent from other researchers. The study involved a number of elements as part of the conceptual framework that were developed by other researchers. Because these elements are not the researcher's original work, permission was sought to include these elements within the study. The two survey instruments, the Teacher Sense of Efficacy Scale (TSES) and the Mathematics Teaching Efficacy Beliefs Instrument (MTEBI) were used to identify teachers' efficacy beliefs. Woolfolk-Hoy, on the Ohio State University website, grants electronic permission to use the TSES (2001) for scholarly research (see Appendix I). Permission was also obtained from Enochs and Huinker to use the MTEBI (2000) (see Appendix J). A key component of the theoretical framework for this study is Bandura's triadic reciprocal determinism diagram (Bandura, 1986). Bandura granted written permission to include the diagram (Appendix B). Lindsey N. Devers, Senior Research Analyst with Learning Sciences International, granted permission to include the Marzano Learning Map (2011) within the dissertation (Appendix C). Its inclusion was necessary to add clarity to the observation scores used as data points within the study. One of the documents, published by the State of Florida to describe the value-added model (Florida Department of Education, 2011) included a notation that it is a draft document, and that it may not be reproduced or cited. The Deputy Commissioner of Education for the State of Florida was contacted and verified that the document is indeed a final document and that it may be cited (Appendix L).

Reporting of data. Results of the study are presented in a manner that prevents subjects or their data from becoming personally identifiable. Due to the use of a surrogate, the researcher

had no identity information for the subjects. Data was presented in an aggregate manner, preventing any individual score or pair of scores from being known to the reader. The manner of reporting meets the ethical standard of privacy (Drew, Hardman, & Hosp, 2008; Howe & Moses, 1999). The analysis methodology for each of the research questions is correlation. Causal attribution is not included as a descriptor within this research, given that correlations indicate relationships between sets of data (Creswell, 2012).

Delimitations

The study is delimited to teachers of fourth- and fifth-grade mathematics in a district on the east coast of Florida. The study is delimited to the Marzano teacher evaluation model, VAM scores in mathematics derived as required by the State of Florida, and evaluation scores for teachers as prescribed by the State of Florida. The researcher also elected to delimit the study to subjects in a single county. Student and staff demographics are only somewhat similar to those of the state, and although efficacy values are not typically related to race or gender (Tschannen-Moran & Hoy, 2007), one should be cautious about generalizing to the state-wide population of teachers.

Limitations

The sample was delimited to 32 teachers, which is sufficient for the correlation model used (Charles & Mertler, 2011; Bartz, 1976), but an increased sample size would render even more reliable results. Additionally, the results are from using only one year of scores for teachers and students, and results could vary if measurements over multiple years were used. All evaluators were trained regarding the Marzano model and demonstrated proficiency in the model utilizing assessments to ensure inter-rater reliability, which acts as a control. Evaluators also participated in annual training and peer review to ensure stable scoring. Limitations of the study

include the ability to generalize to other evaluation models, the ability to generalize to other measures of student growth, and the ability to generalize to other content areas because this study is mathematics specific. Although the Florida model is based on a model founded in Tennessee (Braun, 2005), it is specific to one state, and results of this study may not be similar to those found in different settings. There also exists psychometric criticism of the value-added model and attribution of teacher effect (Baker, Oluwole, & Green, 2013; Braun, 2005; McCaffrey et al., 2003).

The ability to generalize to geographic or demographic situations beyond the state is limited. Only two measures of self-efficacy were used, but other methods of delineating or measuring teacher self-efficacy levels exist, and using other methods could produce different results.

Within the limitations, however, this study does add to the βοδψ of research regarding self-efficacy and teacher classroom performance as well as student academic growth in mathematics. With newly created teacher evaluation systems and an increasing political interest in having value-added measures, this study may be helpful to those looking for antecedents to performance.

A final limitation was the *ceiling effect* of some of the scores. The Marzano observation values, summative evaluation scores, MTEBI scores, and TSES scores were all near the top of their respective scales. Therefore, due to lack of variance, some relationships that may be found with increased variance may not be identified within this study.

Conclusion

Although each of the three questions involved teacher self-efficacy as a predictor variable, the methods used and resultant variables differed. The approach yielded several

important pieces of information: whether teacher self-efficacy, both generally and in mathematics, correlated with classroom observation scores and whether teacher self-efficacy, both generally and in mathematics, correlated with value-added scores. Also, results were able to show the degree to which general teacher self-efficacy, teacher self-efficacy for mathematics, years of teaching, and mathematical training correlated with teachers' evaluation results.

Given that self-efficacy has rarely been used to predict evaluation ratings for teachers, and that the implementation of the Marzano evaluation system and value-added models are relatively recent, the study is timely in adding to the body of research regarding teacher self-efficacy beliefs and the body of research for the Marzano model and value-added methods.

Chapter IV

Results

Introduction

Characteristics of the 32 study subjects are described, including their similarities and differences with district and state teachers. Validity measurements for the MTEBI and TSES are provided. Descriptive statistics are provided for each of the study variables, and each study question is addressed. These results provide a comprehensive report of the quantitative results for the study.

Characteristics of Subjects

Of the 110 fourth- and fifth-grade teachers in the sample school district, 36 elected to participate in the survey, but 43 actually began the survey. Of those who began the survey, 32 provided the data necessary to allow a link to their evaluation results. Of the 32 teachers, two were first year teachers with no evaluation results, and one was new to the school district with no evaluation results. Three of the teachers who had been members of the district staff also did not actually teach mathematics during the previous year, so they did not have a mathematics VAM score but did have a summative evaluation score.

The mean age of participants was 42.06 years. The minimum age was 21 and the maximum was 62. The distribution pattern appears to be normal, based on visual inspection of the frequency histogram of subject ages (see Figure 3). Analysis of skewness and kurtosis yielded respective values of .074 (*SE* = .414) and -.337 (*SE* = .809), further verifying a normal distribution.

Figure 3

Age of Teachers
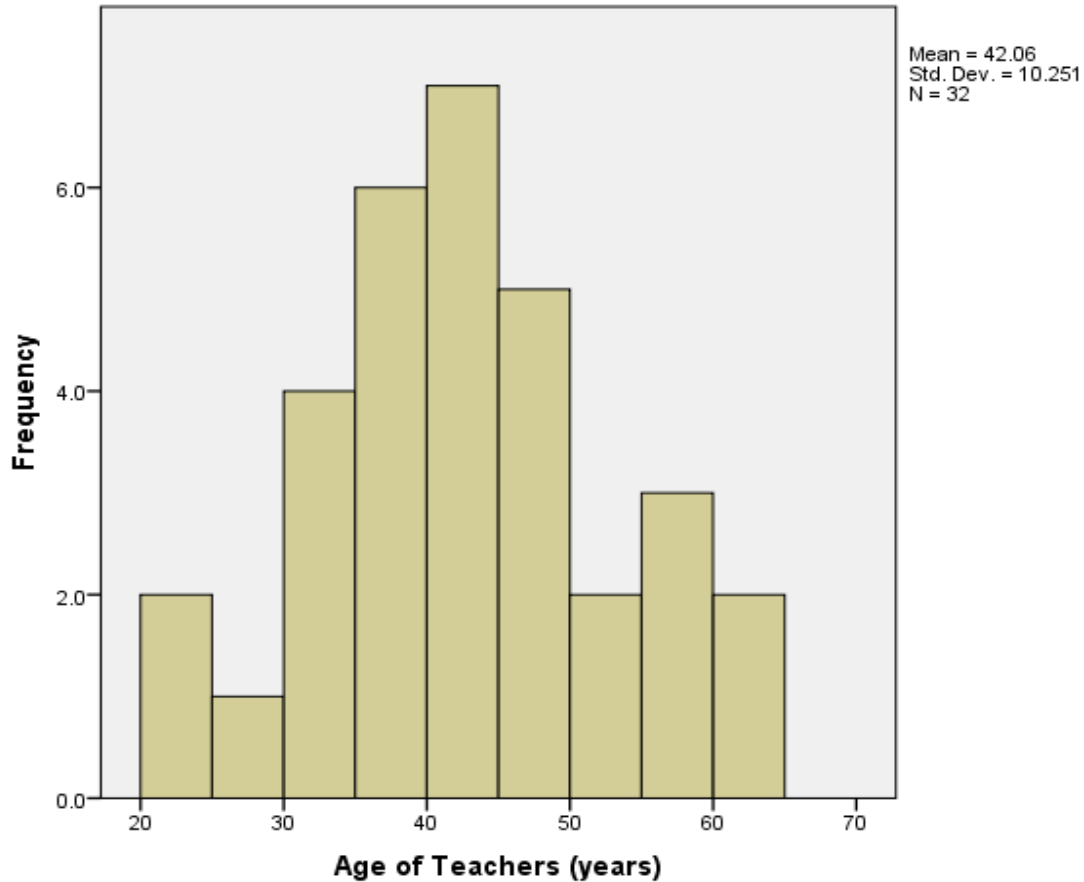
Note. This histogram reflects the distribution of subjects' reported ages.

The mean for participants' years of teaching was 15.25 years. The minimum was 1 and the maximum was 34. The distribution pattern appears to be normal based on visual inspection of the frequency histogram of subjects' years of teaching (see Figure 4). Analysis of skewness and kurtosis yielded respective values of .295 (SE = .414) and -.056 (*SE* = .809), further verifying a normal distribution.

Figure 4

Years of Teaching

Note. This histogram reflects the distribution of subjects' reported years of teaching.

Mean = 15.25
Std. Dev. = 8.144
N = 32
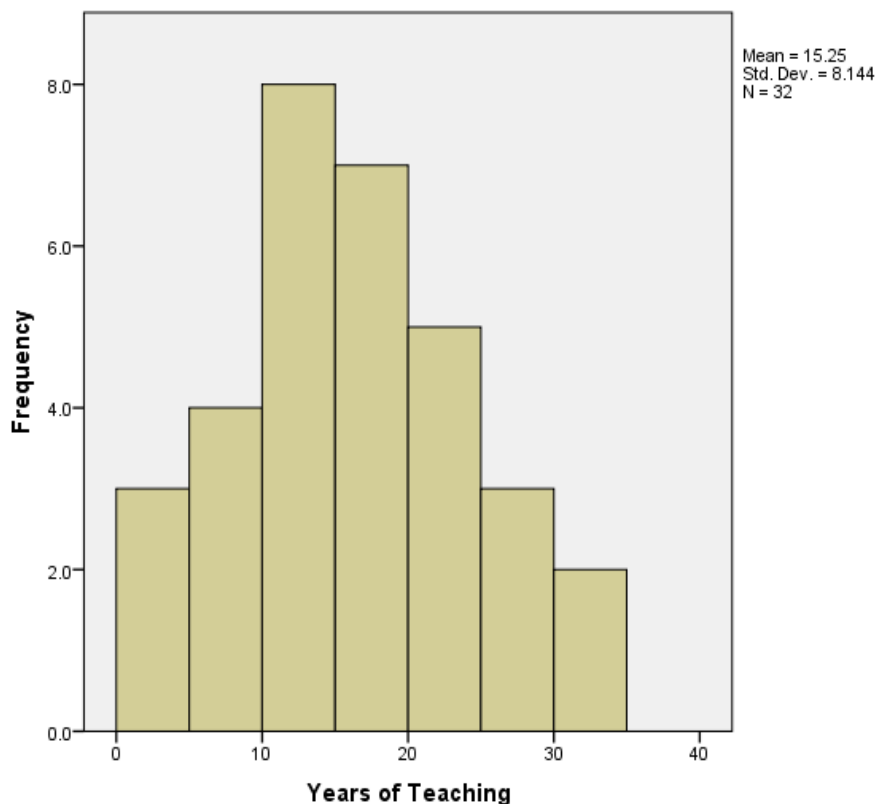
The racial/ethnic composition of the 32 teachers in the sample was 93.8% White ($n = 30$),

3.1% Black ($n = 1$), and 3.1% Hispanic/Latino ($n = 1$). Initially, the teachers' data was compared

with the district's and state's teacher demographics by using a chi-square goodness of fit test.

However, the limited sample size caused a failed assumption necessary for implementation of a

chi-square, namely that all expected values would be greater than five. For small samples, an

alternate method to compare a sample's composition to a population is the Fisher's exact test

(Lantz, 1978). The sample's racial demographics were statistically similar to those of the

district's personnel (p = .49085, α < .05). However, the sample's racial profile was significantly

different from that of personnel from the state of Florida (p = .00352, not significant, α < .01).

Because of this difference, a comparison between the sample's district and the state of Florida

was conducted using a chi-square goodness of fit test. The racial demographics of the sample's

district differed significantly from the state demographics (p = .000, α < .01). Further details

regarding this analysis are available in Appendix A. The *Simple Interactive Statistical Analysis*

web tool was used to conduct the Fisher's exact test (Uitenbroek, 1997). Whether the sample

school district demographics and state demographics are similar is likely inconsequential

because self-efficacy has been shown to not vary significantly based on teacher race (Tschannen-

Moran & Hoy, 2007).

The gender distribution of the 32 teachers in the sample who provided gender

information was 12.5% male (n = 4) and 87.5% female (n = 28). A chi-square goodness of fit

test showed that the sample's gender demographics were statistically similar to those of the

district's personnel and the state's personnel (both α < .05). According to a chi-square analysis,

the district gender demographics, however, did vary significantly from the state profile (α < .01).

Details regarding the gender demographic tests are also available in Appendix A.

Validity of the Teachers' Sense of Efficacy Scale and the Mathematics Teacher Efficacy

Beliefs Instrument

Validity was tested for this implementation of both the Teachers' Sense of Efficacy Scale

(Tschannen-Moran & Hoy, 2001) and the Mathematics Teaching Efficacy Beliefs Instrument

(Enochs et al., 2000). Because both instruments are commonly used for teacher self-efficacy

research and validity is generally accepted (see Table 2), a confirmatory factor analysis was

conducted to verify whether this administration of the surveys was similar to those of other

researchers. As expected, Cronbach's alpha levels were consistent with previous research and sufficiently reliable (see Table 3 and Table 4).

The only Cronbach's alpha level that was less than the generally accepted threshold of $\alpha = .7$ was the outcome expectancy subscale of the MTEBI. Further exploration of this subscale finding showed that if question nine was removed, the alpha level increased from .654 to .693. The survey item is, "The inadequacy of a student's mathematics background can be overcome by good teaching." On face validity, this item appears to be a proper descriptor of outcome expectancy. To remove the item would mean implementation of the survey in a form different than its consistently tested format. Outcome expectancy subscale scores are frequently lower than the personal mathematics teaching efficacy and MTEBI general alpha levels (see Table 2), so this implementation is consistent with previous findings. This item also did not compromise the MTEBI general reliability level. For these reasons, the item was not excluded when data analyses were conducted regarding the research questions.

Table 3

Validity Measures for the Teachers' Sense of Efficacy Scale (TSES)

Overall Validity	Elements' Factor Validities		
	Instruction	Management	Engagement
.848	.758	.766	.811

Note. Validity scores are Cronbach's alpha measures.

Table 4

Validity Measures for the Mathematics Teaching Efficacy Beliefs Instrument (MTEBI)

Overall Validity	Elements' Factor Validities	
	Outcome Expectancy	Personal Mathematics Teaching Efficacy
.783	.654	.899

Note. Validity scores are Cronbach's alpha measures.

Descriptive Results for Survey Variables

Descriptive statistics for each variable in the study were reviewed. The results of these analyses are described in this section.

College Coursework. A composite value was computed to identify the relative rigor of college mathematics coursework for each subject. Each course was assigned a numerical value identifying its relative difficulty. A composite mean score was then computed to denote a holistic description of the mathematical coursework each had successfully passed. The numerical values were as follows:

Remedial College Mathematics 1

Mathematics Methods 1

College Algebra 2

College Geometry 2

Trigonometry 2

Math Analysis or Pre-Calculus 3

Statistics 3

Calculus 4

Discrete Mathematics 4

Use of this system allowed for a relative ranking of the teachers' math experience, with a "1" denoting limited mathematical rigor and a "4" denoting maximum mathematical rigor. The subjects' math attainment scores ranged from 1.00 to 2.67, with a mean of 1.96 ($SE = .079$). The Shapiro-Wilk test confirmed normality of college educational attainment in mathematics (.950, $p > .05$).

Professional Development. Teachers self-reported the number of mathematics professional development units in which they had participated within the previous five years. Subjects were asked to identify the number of units by choosing 0, 1-5, 6-10, 11-15, 16-20, or 21+ units. Because the choices included no professional development and greater than 21 units, which have different spans than the other options, the ratings were ranked and treated as ordinal variables. Subjects noted participation in all six levels of professional development (see Figure 5). However, the frequencies were not statistically normal. The Shapiro-Wilk test yielded a score of .893 ($p < .05$). This finding lends itself to use of a Spearman rho when identifying correlations, given that assumptions of normality and interval scale were not met.

Figure 5

Frequency of Professional Development in Mathematics
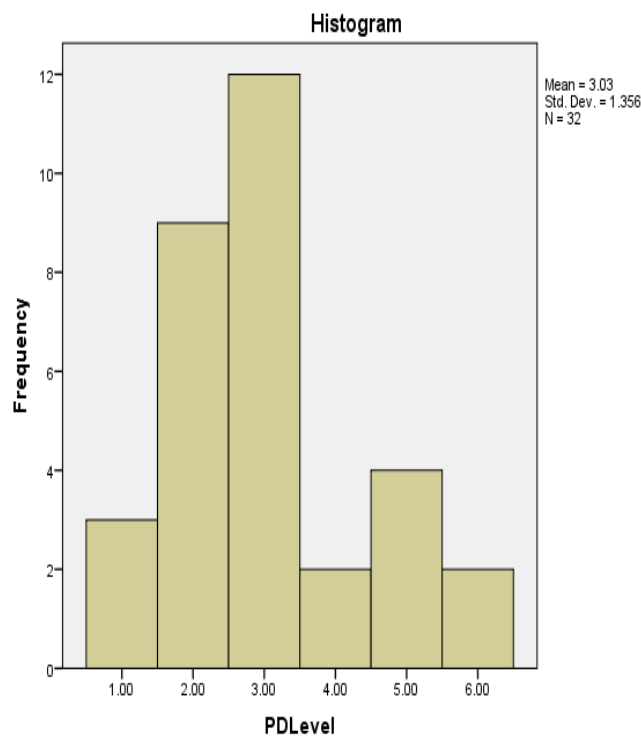
Note. Subjects are ranked by level, depending on the number of hours of mathematics instructional professional development they have participated in within the last five years. A rank of 1.00 represents zero hours, 2.00 represents 1-5 hours, 3.00 represents 6-10 hours, 4.00 represents 11-15 hours, 5.00 represents 16-20 hours, and 6.00 represents 21+ hours.

MTEBI general. The MTEBI general scores yielded a mean of 4.038 (*SE* = .0598). The minimum score was 3.39 and the maximum score was 4.65. The data met assumptions of normality with kurtosis of -.581 (*SE* = .809) and skewness of -.403 (*SE* = .414). A Shapiro-Wilk test also yielded a value of .977 (*p* > .05), further verifying normality. Figure 6 shows the distribution.

Figure 6

Distribution of MTEBI General Scores
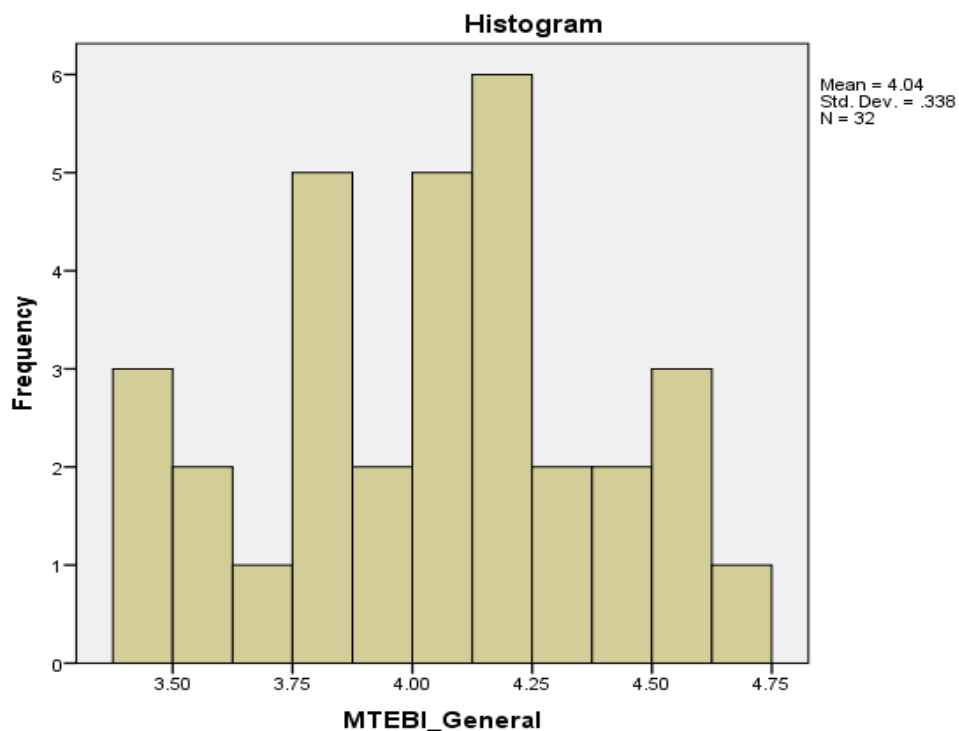
MTEBI subscales. The MTEBI has two subscales: outcome expectancy and personal mathematics teaching efficacy. The outcome expectancy subscale yielded a mean value of 3.63 for this administration (*SD* = .414, *N* = 32). The minimum score was 2.60, and the maximum score was 4.20. The values for skewness and kurtosis were -.479 (*SE* = .414) and -.349 (*SE* =

.809), respectively. A Shapiro-Wilk test yielded a value of .950 ($p > .05$), verifying the

assumption of normality (see Figure 7).

Figure 7

Frequencies of MTEBI Outcome Expectancy Subscale Values
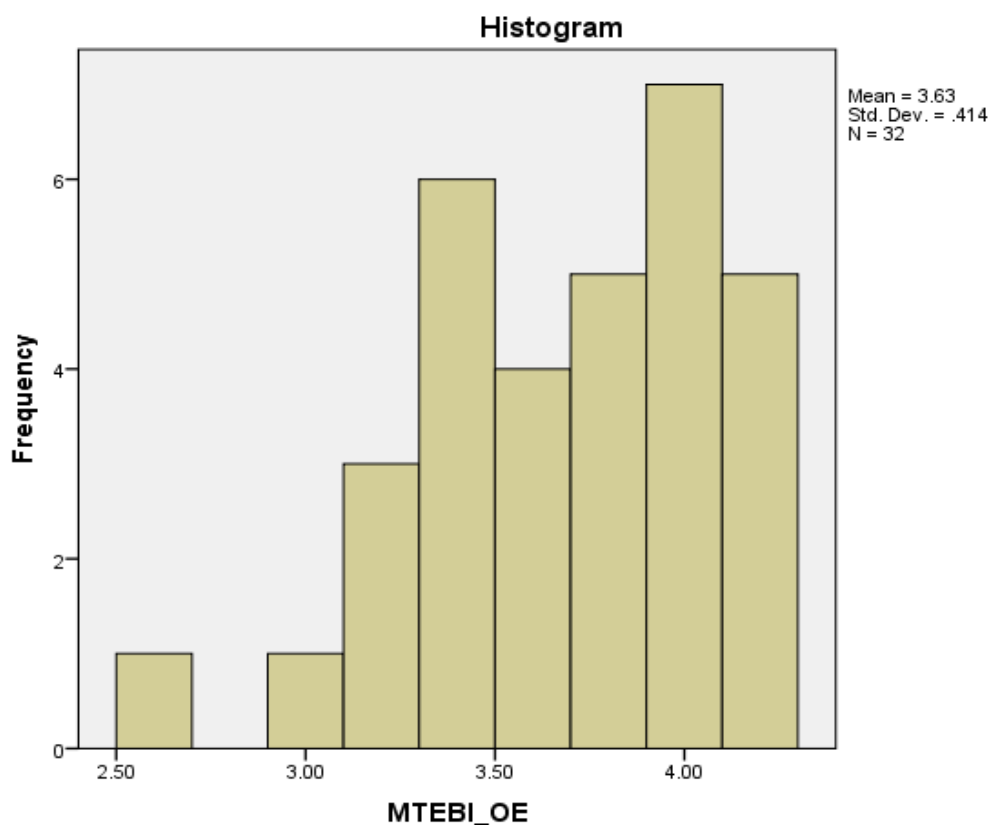
The personal mathematics teaching efficacy subscale yielded a mean value of 4.354 (*SD* = .543, *N* = 32). The minimum score was 3.08 and the maximum score was 5.00. Testing for skewness yielded a value of -.580 (*SE* = .414) and kurtosis yielded a value of -.480 (*SE* = .809). Although these values are consistent with an assumption of normality, the Shapiro-Wilk test yielded a value of .927 ($p < .05$), failing to meet the $\alpha = .05$ level required to verify normality. By viewing the histogram (Figure 8), one can see that the values were clustered toward the high portion of the scale, which is probably the reason the scores cannot be considered normal. This

subscale is not a primary element of the study's research questions, but the lack of normality is worth noting.

Figure 8

Frequencies of MTEBI Personal Mathematics Teaching Efficacy Subscale Values
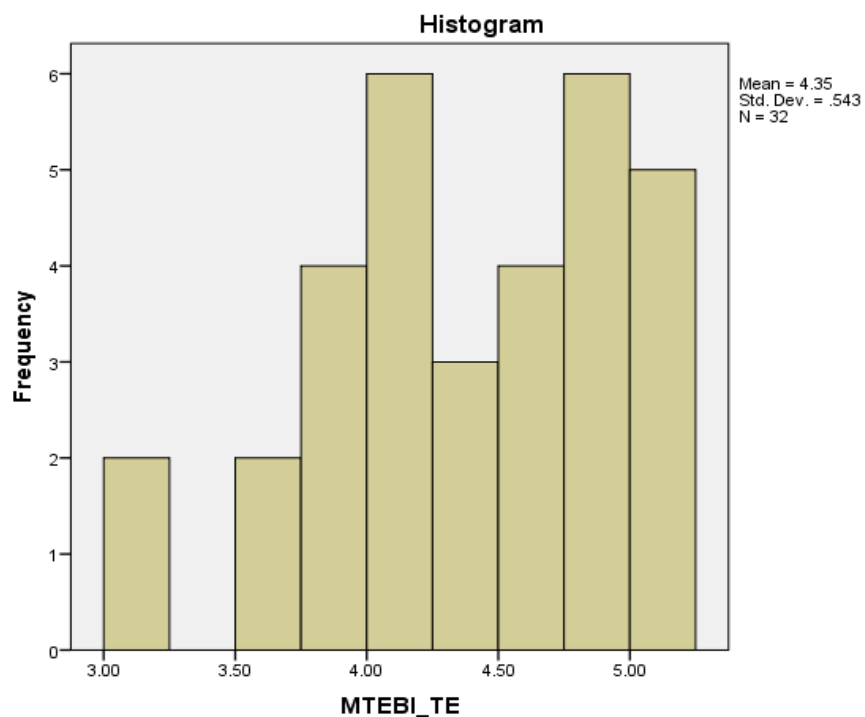
TSES general. The TSES general scores yielded a mean of 7.849 (*SD* = .120). The minimum score was 6.42 and the maximum score was 9.00. The data met assumptions of normality with kurtosis of -.494 (*SE* = .809) and skewness of -.389 (*SE* = .414). The Shapiro-Wilk test also yielded a value of .959 (*p* > .05), further verifying normality. Figure 9 shows a histogram of the results.

Figure 9

Frequencies of TSES General Scores
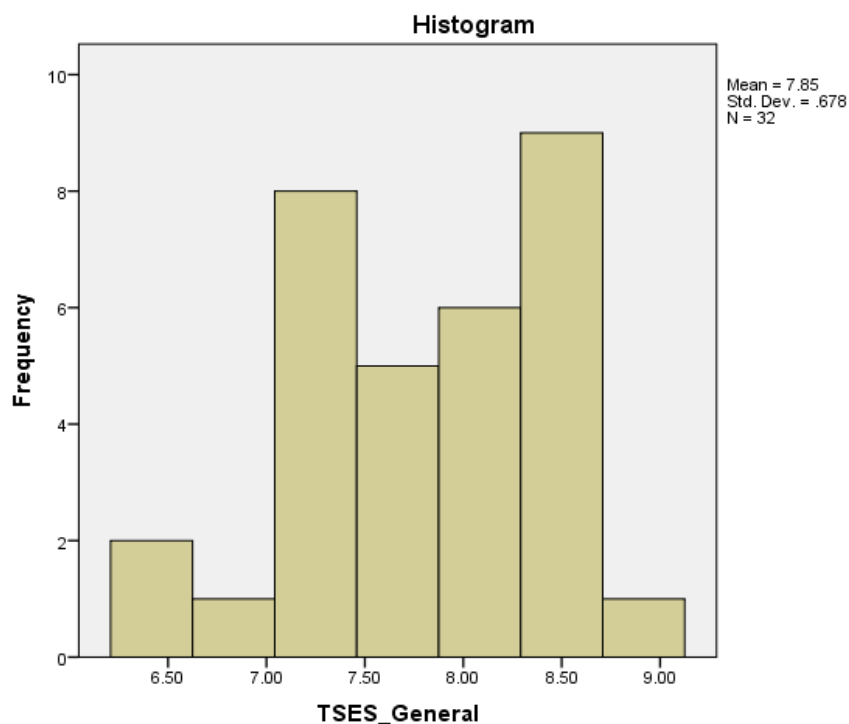
TSES subscales. The TSES has three subscales: engagement, instruction, and

management. The engagement subscale (see Figure 10) yielded a mean value of 7.648 (*SD* =

.168, *N* = 32). The minimum score was 5.25, and the maximum score was 9.00. The values for

skewness and kurtosis were -.823 (*SE* = .414) and -.320 (*SE* = .809), respectively. Although

these values are consistent with an assumption of normality, A Shapiro-Wilk test yielded a value

of .891 (*p* < .05), failing to meet the α = .05 level required to verify normality. Although the

subscale is not directly related to any of the study's research questions, the lack of normality

should be noted.

Figure 10

Frequencies of TSES Engagement subscale Scores
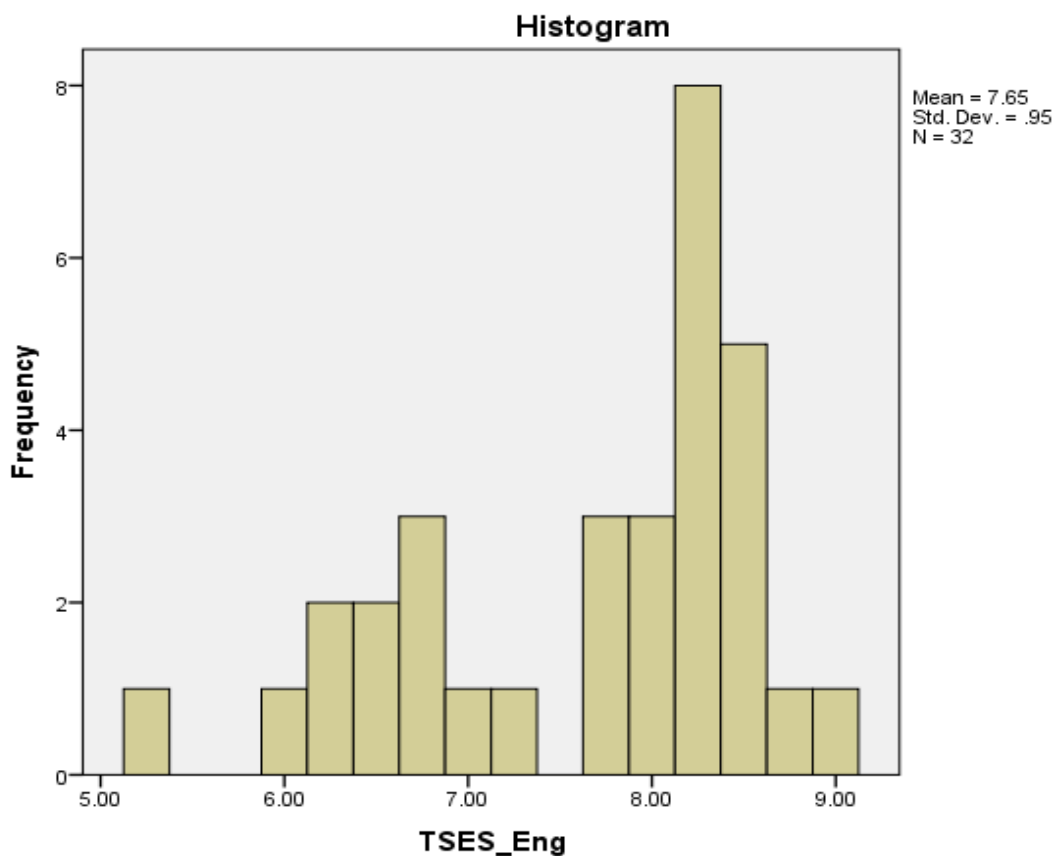
The instruction subscale of the TSES (see Figure 11) yielded a mean value of 7.758 (*SD* = .1570, *N* = 32). The minimum score was 5.25 and the maximum score was 9.00. Testing for skewness yielded a value of -.567 (*SE* = .414) and kurtosis yielded a value of .518 (*SE* = .809). These values are consistent with an assumption of normality, and the Shapiro-Wilk test yielded a value of .945 (*p* > .05), meeting the $\alpha = .05$ level required to verify normality. The instruction subscale results can be considered normal.

Figure 11

Frequencies of TSES Instructional subscale Values
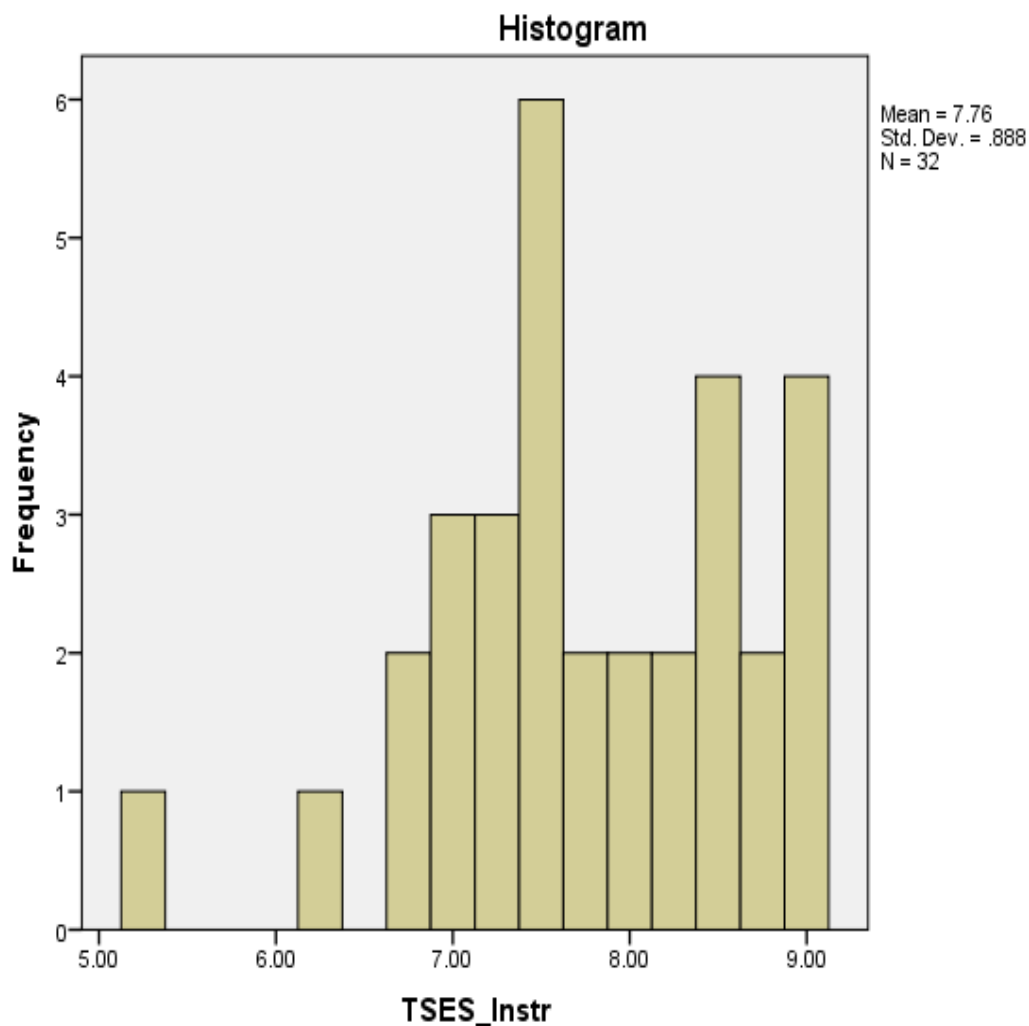
The classroom management subscale (see Figure 12) yielded a mean value of 8.141 (*SD* = .1289,

N = 32). The minimum score was 6.75, and the maximum score was 9.00. The values for

skewness and kurtosis were -.301 (*SE* = .414) and -1.086 (*SE* = .809), respectively. Although

these values are consistent with an assumption of normality, a Shapiro-Wilk test yielded a value

of .901 (*p* < .05), failing to meet the α = .05 level required to verify normality. Although the

subscale is not directly related to any of the study's research questions, the lack of normality should be noted.

Figure 12

Frequencies of TSES Classroom Management Subscale Values
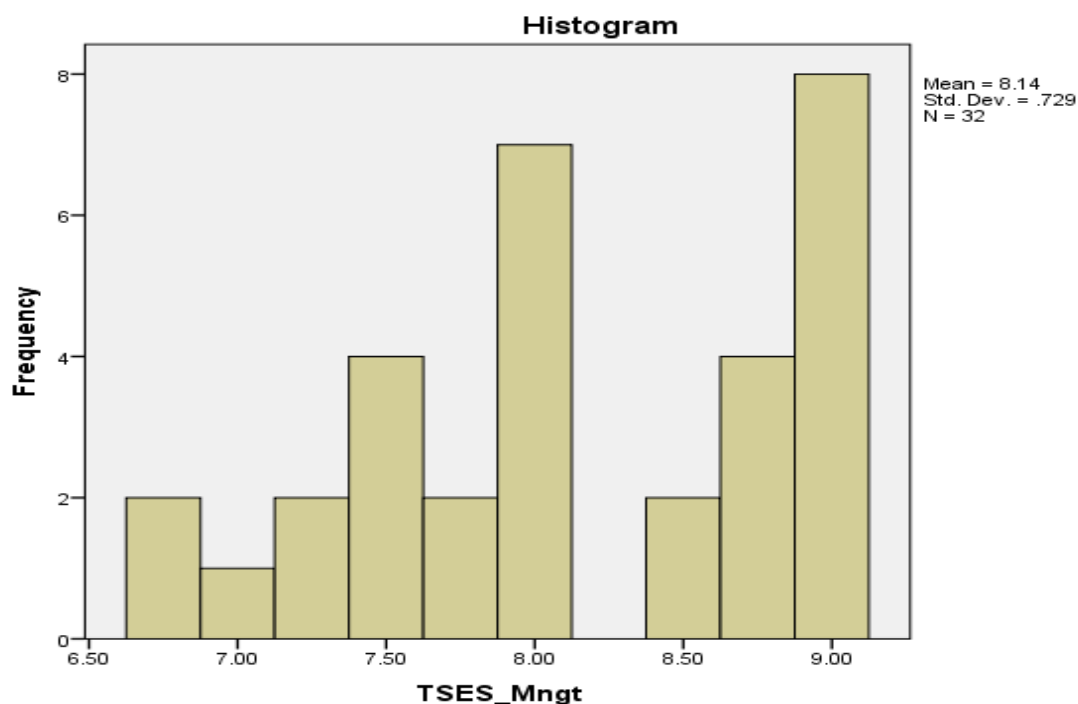
Mathematics value-added scores. For the subjects of this study, the mean mathematics VAM score was .104 ($SD = .171$, $n = 26$). The minimum VAM value was -.206 and the maximum value was .524. The values for skewness and kurtosis were .621 ($SE = .456$) and .333 ($SE = .887$), respectively. The Shapiro-Wilk test yielded a value of .966 ($p > .05$). These results lead to an appropriate assumption of normality. This can also be verified by viewing the frequency histogram (Figure 13). VAM scores are derived from z-scores across the state of Florida, so one would expect the subjects' scores to follow a normal pattern.

Figure 13

Frequencies of Mathematics Value-Added Measures
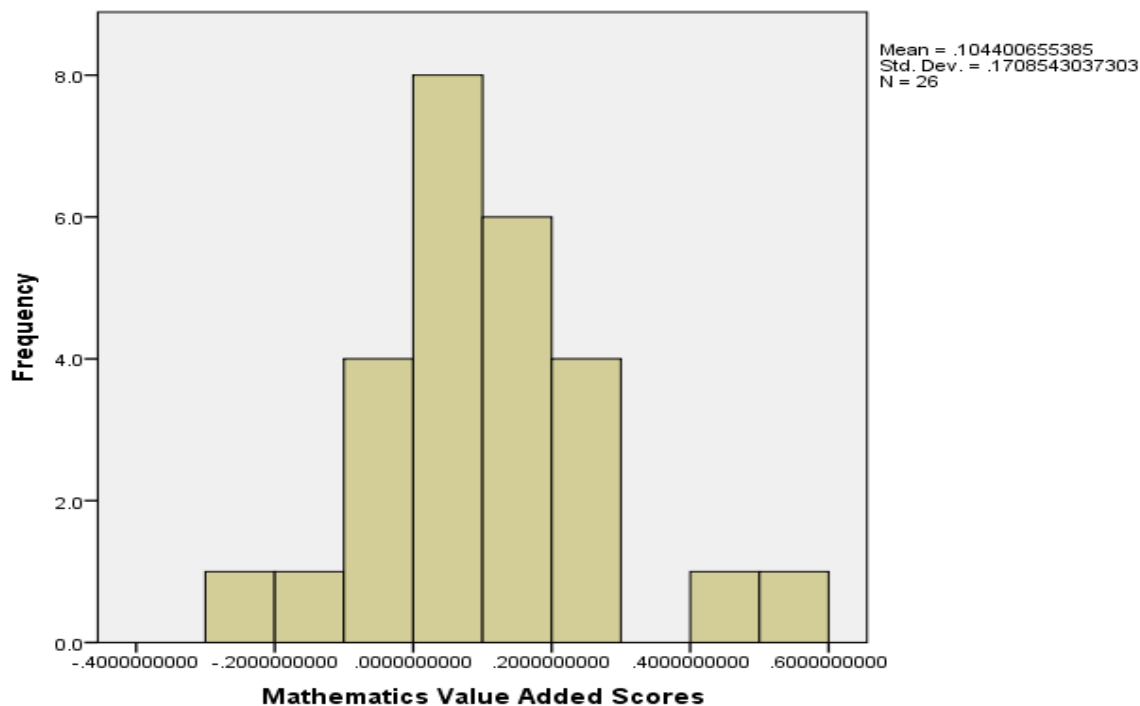
Note. Mathematics value-added scores follow a reasonably normal pattern. Note, however that the "zero" point is slightly to the left of the peak of the curve.

Marzano instructional practice scores. All subjects had earned Marzano evaluation scores of Needs Improvement, Effective, or Highly Effective. One subject had a score of Needs Improvement (denoted by "2"), seventeen subjects had earned an Effective score (denoted by "3"), and nine subjects had earned a score of Highly Effective (denoted by "4"). Given the lack of variance for these scores, one would expect the data to fail tests for normality. The Shapiro-Wilk value was .604 ($p < .05$), verifying the expected non-normal distribution (see Figure 14).

Figure 14

Frequencies of Marzano Instructional Practice Scores
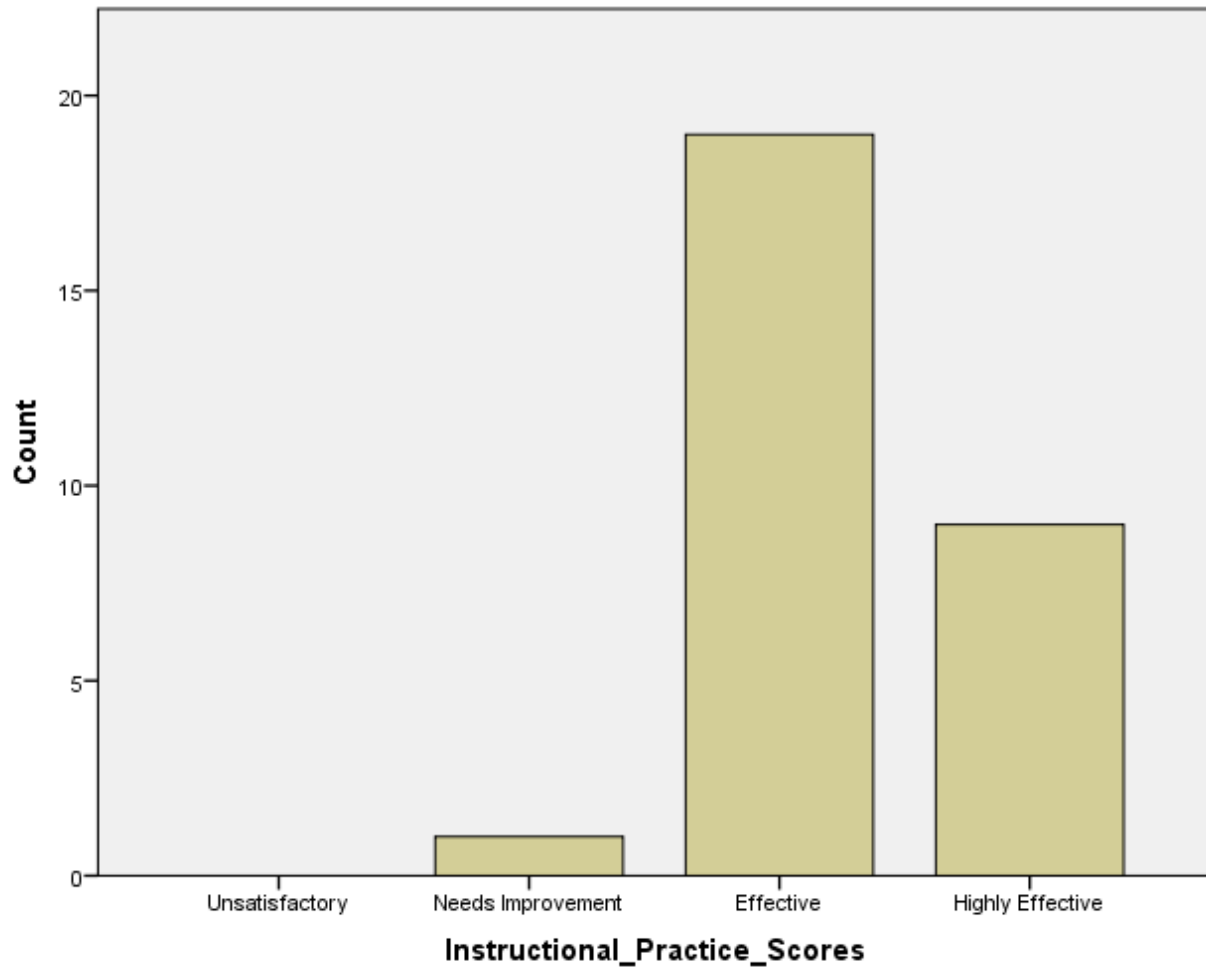
Summative evaluation scores. The mean summative evaluation score for the teacher-subjects was 3.35 (*SE* = .095, *n* = 29). The minimum score was 2.00 and the maximum score was 4.00. The lack of score variance leads one to suspect that the distribution was not normal. The Shapiro-Wilk test yielded a value of .914 (*p* < .05) verifying that the distribution was not normal (see Figure 15).

Figure 15

Frequencies of Summative Evaluation Scores
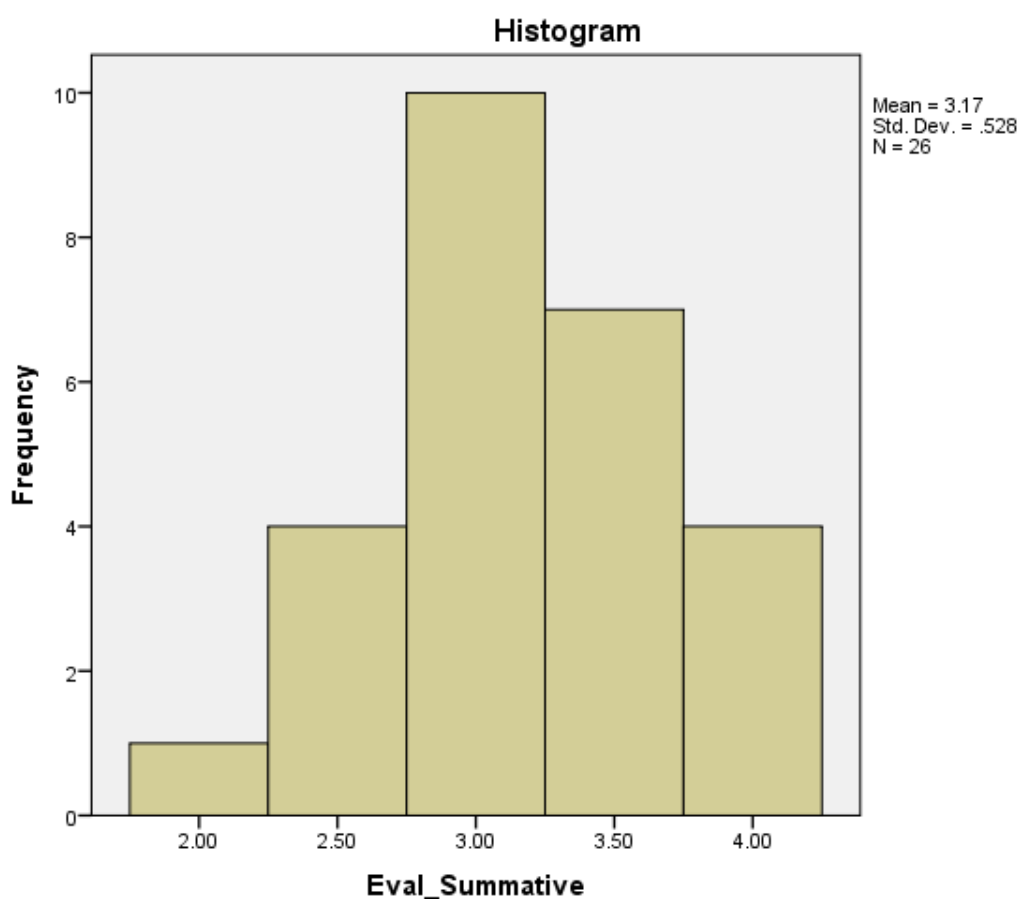
Note. Shows the number of subjects with each summative evaluation score. Scores are composed of 40% value-added scores and 60% Marzano instructional practice scores.

The lack of normality for the Marzano observation scores and summative evaluation scores verify that the appropriate correlational test is a Spearman rho. The Spearman rho is appropriate when normality cannot be assumed, or when one or both variables are ordinal (Field, 2013).

Research Questions

Question 1. What is the relationship between teachers' levels of self-efficacy and teacher observation scores derived from the Marzano evaluation system? To address this

question, both mathematics teaching self-efficacy and general teaching self-efficacy were explored as possible predictor variables.

Mathematics self-efficacy and Marzano scores. The analysis conducted was to identify the degree of relationship between the subjects' mathematics self-efficacy scores, as measured by the MTEBI, and their teacher instructional practice (IP) scores derived from the Marzano evaluation system. To conduct this analysis, a Spearman rho correlation was run. The null hypothesis was that there is no significant relationship between MTEBI scores and IP scores. The alternative hypothesis was that there is a significant relationship between MTEBI scores and IP scores.

Based on the Spearman rho results (see Table 5), the null hypothesis is retained, leading to the conclusion that there is not a significant relationship between teachers' general self-efficacy for teaching mathematics and teachers' scores for instructional practice. The low coefficient, even if the correlation had shown to be significant, would verify no relationship due to its proximity to zero.

Table 5

Correlations between Mathematics Efficacy Levels and Marzano Instructional Practice Scores

	IP scores		
	r	Significance	*n*
MTEBI	.061	.755	29
Outcome expectancy factor	.261	.171	29
Personal mathematics teaching efficacy factor	-.090	.641	29

Note: No correlations were significant ($\alpha < .05$)

Although not a primary focus of the research question, subscale analyses were also conducted. The two subscale factors for the MTEBI are outcome expectancy and personal

mathematics teaching efficacy (see Table 5).Although the correlation coefficient was somewhat

higher for the outcome expectancy factor than for the overall mathematics self-efficacy score, the

significance level was not sufficient to verify a relationship. In addition, the correlation

coefficient for the personal mathematics teaching efficacy factor was close to zero and slightly

negative. Therefore, the null hypothesis is retained of no relationship between the subscale

factors and IP scores.

      General self-efficacy and Marzano scores. The other portion of question 1 attempts to

identify the degree of relationship between the subjects' general teaching self-efficacy scores, as

measured by the TSES, and teacher IP scores derived from the Marzano evaluation system. To

conduct this analysis, a Spearman rho correlation was run. The null hypothesis was that there is

no significant relationship between TSES scores and IP scores. The alternative hypothesis was

that there is a significant relationship between TSES scores and IP scores.

      Based on the Spearman rho results (see Table 6), the null hypothesis is retained, leading

to the conclusion that there is not a significant relationship between TSES scores and IP scores

for the subjects. The low negative coefficient, even if the correlation had shown to be significant,

would actually verify no relationship between general teacher efficacy and Marzano IP scores.

Table 6

Correlations between Teacher Self-Efficacy Levels and Marzano Instructional Practice Scores

	IP scores		
	r	Significance	*n*
TSES	-.074	.701	29
Engagement factor	.149	.441	29
Instructional factor	-.195	.311	29
Management factor	.048	.806	29

Note: No correlations were significant ($\alpha < .05$)

Although not a primary subject of this question, TSES subscale analyses were also conducted. The three subscale factors for the TSES instrument are efficacy for engagement, efficacy for instruction, and efficacy for management (see Table 6). Although the correlation coefficient was somewhat higher for the engagement factor than for the overall teacher self-efficacy score, the significance level was not sufficient to verify a relationship. For the instruction factor, the negative correlation coefficient would have supported an inverse relationship between general teacher self-efficacy and IP scores, had the test been significant. For the final factor of management, the coefficient was close to zero so that even if significance had existed, no relationship would have been indicated. In summary, the null hypothesis of no relationship to IP scores is retained for all three TSES subscale factors.

Further analyses. As an additional step, even though it was not part of the data analysis plan, a *t* test for equality of means was conducted. This extra step was taken because the IP data yielded subjects who had either Effective or Highly Effective scores, with the exception of one subject who was assigned a Needs Improvement score. None of the subjects who completed the survey were categorized as Unsatisfactory. The lack of variance for the IP scores may have contributed to the lack of relationship between IP performance and self-efficacy. For each of the *t* tests, the null hypothesis was that there is no significant difference between the means. The alternative hypothesis for each case was that there is a significant difference between the means. For each test, equal variances were assumed due to *F* scores yielded from Levene's test for equality of variances ($\alpha > .05$).

For the MTEBI *t* test analysis, the subjects were divided into two groups, those with Effective scores, and those with Highly Effective scores. For the general MTEBI measure,

subjects with Effective evaluations did not show significantly different levels than teachers with Highly Effective evaluations. The null hypothesis was retained, indicating no significant difference (see Table 7).

Table 7

Mathematics Teaching Efficacy t test Results

	Effective Evaluation			Highly Effective Evaluation			*t* scores	
	M	*SE*	*n*	*M*	*SE*	*n*	*t* score	*Sig.*
MTEBI	4.0815	.06547	19	4.1227	.12746	9	-3.20	.752
Outcome Expectancy Factor	3.5474	.09532	19	3.8333	.12134	9	-1.765	.089
Personal Mathematics Teaching Efficacy Factor	4.4960	.12106	19	4.3469	.12106	9	.726	.475

Note: No differences demonstrated significance ($\alpha < .05$, two-tailed)

For the MTEBI outcome expectancy subscale *t* test analysis, subjects with Effective evaluations also did not show significantly different levels than teachers with Highly Effective evaluations. The null hypothesis is retained, indicating no significant difference (see Table 7).

For the MTEBI personal mathematics teaching efficacy subscale *t* test analysis, subjects with Effective evaluations again did not show significantly different levels than teachers with Highly Effective evaluations. In fact, although not significant, the Highly Effective teachers' mean was slightly lower than the mean for Effective teachers. The null hypothesis is retained, indicating no significant difference (see Table 7).

For the TSES *t* test analysis, the subjects were divided into the same two groups: those with Effective scores, and those with Highly Effective scores. For the general TSES measure, the mean self-efficacy score for Highly Effective teachers was slightly lower than the score for

Effective teachers. However, the difference is not significant and the null hypothesis was retained (see Table 8).

For the TSES engagement, instructional, and management subscale *t* test analyses, subjects with Effective evaluations did not show significantly different levels than teachers with Highly Effective evaluations. For each subscale, the null hypothesis is retained (see Table 8).

Table 8

General Teacher Efficacy t test Results

	Effective Evaluation			Highly Effective Evaluation			*t* scores	
	M	*SE*	*n*	*M*	*SE*	*n*	*t* score	*Sig.*
TSES	7.9605	.13524	19	7.8056	.27604	9	.570	.574
Engagement Factor	7.6447	.18762	19	7.8056	.36271	9	-.437	.666
Instructional Factor	7.9342	.15513	19	7.4167	.39747	9	1.473	.153
Management Factor	8.3026	.14632	19	8.1944	.26607	9	.387	.702

Note: No differences demonstrated significance ($\alpha < .05$, two-tailed)

In summary, for question 1, no significant correlations were found between Marzano instructional practice scores and teachers' sense of efficacy, either generally or specifically in mathematics. Because there were no significant relationships, the planned point-biserial regression was not conducted.

Given a lack of variance in instructional practice scores, an analysis was conducted to identify any significant differences between efficacy levels of teachers who had Effective scores and those who had Highly Effective scores. No significant differences were found for either the general measures (TSES), mathematics-specific measures (MTEBI), or their respective subscales.

Question 2. What is the relationship between teachers' sense of self-efficacy and their Florida State value-added scores in mathematics? To address this question, both mathematics teaching self-efficacy and general teaching self-efficacy were explored as possible predictor variables.

Mathematics self-efficacy and mathematics value-added scores. The analysis was conducted to identify the degree of relationship between the subjects' mathematics self-efficacy scores, as measured by the MTEBI, and teacher value-added scores in mathematics. To conduct this analysis, a Pearson correlation was run. The null hypothesis was that there is no significant relationship between MTEBI scores and value-added scores. The alternative hypothesis was that there is a significant relationship between MTEBI scores and value-added scores.

Initially, correlations were run between MTEBI scores and mathematics VAM scores. Both are scale variables, so the Pearson model was selected. Results are shown in Table 9.

Table 9

Correlations for Mathematics Teaching Efficacy and Mathematics VAM Scores

	Mathematics VAM scores		
	r	*Significance*	*N*
MTEBI	.527	.006*	26
Outcome expectancy factor	.582	.002*	26
Personal mathematics teaching efficacy factor	.264	.192	26

Note: * denotes significance ($\alpha < .01$)

The resulting r^2 value is .278, with an adjusted r^2 value of .248. The resultant effect size based on the adjusted r^2, therefore, is .3298, which is considered a large effect size (Cohen, 1992; Durlak, 2009). These data results support the concept that mathematics VAM can be predicted by teachers' mathematics self-efficacy.

The model's residuals were visually inspected for heteroscedasticity and did not show a problematic pattern.When a linear regression was run for the MTEBI scores and mathematics VAM scores, the unstandardized coefficient was .285 with an intercept of -1.869. The linear regression equation, therefore, is:

$$math\ VAM = -1.869 + .285 * MTEBI$$

Each of the MTEBI subscales was also analyzed for a relationship with mathematics VAM scores (see Table 9). For outcome expectancy, the alternative hypothesis was adopted, indicating that outcome expectancy did significantly predict mathematics VAM. However, for the personal mathematics teaching efficacy subscale, the null hypothesis was supported; there exists no significant relationship for this subscale.

General self-efficacy and mathematics value-added scores. The analysis was conducted to identify the degree of relationship between the subjects' general self-efficacy scores, as measured by the TSES, and teacher value-added scores in mathematics. To conduct this analysis, a Pearson correlation was run. The null hypothesis was that there is no significant relationship between TSES scores and value-added scores. The alternative hypothesis was that there is a significant relationship between TSES scores and value-added scores.

Initially, correlations were run between TSES scores and mathematics VAM scores. Both are scale variables, so the Pearson model was selected. Based on the Pearson results (see Table 10), the null hypothesis was supported that there is no significant relationship between general teachers' self-efficacy and their mathematics VAM scores.

The three subscales of the TSES were also analyzed to identify relationships with mathematics VAM scores. The subscales are engagement, instruction, and management (see

Table 10). The null hypothesis was supported that there is no significant relationship between teachers' engagement self-efficacy and their mathematics VAM scores.

For TSES-instruction and VAM, the null hypothesis was also supported. Even had the finding been significant, the correlation value was close to zero. There is no significant relationship between teachers' engagement self-efficacy and their mathematics VAM scores (see Table 10).

For TSES-management and VAM, the null hypothesis was again supported. Even had the finding been significant, the correlation value was again close to zero. There is no significant relationship between teachers' management self-efficacy and their student's mathematics VAM scores (see Table 10).

Table 10

*Correlations for Teachers' Sense of Effic*acy Scale *and Mathematics VAM Scores*

	Mathematics VAM scores		
	r	*Significance*	*n*
TSES	.172	.402	26
Engagement factor	.313	.119	26
Instructional factor	.056	.787	26
Management factor	.019	.928	26

Note: No correlations were significant ($\alpha < .05$)

In summary, the answers to question 2 supported by the study data are that mathematics VAM can be predicted by teachers' mathematics self-efficacy, with a large effect size (Cohen, 1992; Durlak, 2009). However, mathematics VAM is not predicted by general self-efficacy. Within the mathematics self-efficacy construct, outcome expectancy significantly predicts

mathematics VAM, but the personal mathematics teaching subscale does not. No subscales of the TSES predict mathematics VAM.

Question 3. What proportion of influence do the following have on a teacher's evaluation rating: teacher sense of efficacy, as measured by the Tschannen-Hoy instrument; mathematics teaching efficacy, as measured by the MTEBI; years of teaching experience; mathematics instruction professional development; and mathematics coursework attained? Question 3 was analyzed using a multiple regression approach. For this question, the summative evaluation score was the dependent variable. The summative score, in accordance with Florida Statute (1012.34), is derived by combining the Marzano IP observation score with value-added scores as 40% of the total and 60% of the total, respectively. For fourth- and fifth-grade teachers, the value-added scores included in the summative evaluation were composed of both reading and mathematics VAM scores. The independent variables considered for the multiple regression included:

- mathematics teacher self-efficacy scores (MTEBI),

- general teacher self-efficacy scores (TSES),

- college mathematics coursework,

- mathematics instruction professional development, and

- years of teaching.

Initially, Pearson and Spearman rho correlations were run to identify significance, as appropriate. One of the independent variables, professional development level, was ordinal, which required a Spearman rho approach. All other variables were scale scores, requiring a Pearson correlation.

When SPSS was used with all five independent variables to define a model, it only included MTEBI (math teaching efficacy). Professional development and MTEBI, however, both

independently correlated significantly with evaluation scores, .416, *p* =.028 and .522, *p* = .002,

respectively, so this was explored further as the linear regression model was developed.

     Professional development and mathematics teaching efficacy were the only real

contributing factors toward summative evaluations. Development of a linear regression model,

therefore, focused on these two elements and disregarded the others. Attempts were made to

develop a hierarchical multiple regression model. The two elements were introduced to the

model in two different orders to explore the relationships and correlations.

Table 11

Correlations for Multiple Regression Elements

	Summative Evaluation Scores * Scores: 1-4, *min* = 2.0, *Max* = 4.0, *M* = 3.1393, *SD* = .52939	
	Pearson *r* Significance	Spearman rho Significance
TSES, general Scores 1-9, *min* = 6.42, *max* = 9.00, *M* = 7.8490, *SD* = .67814	0.190 .167	
MTEBI, general Scores 1-5, *min* = 3.39, *max* = 4.65, *M* = 4.0381, *SD* = .33834	0.522 ** .002	
Years of teaching *min* = 1, *max* = 34, *M* = 15.25, *SD* = 8.144	0.120 .272	
Professional development level* Scores 1-6, *min* = 1, *max* = 6 *M* = 3.0313, *SD* = 1.0313		0.416 ** .028
College mathematics coursework Scores 1-4, *min* = 1.00, *max* = 2.67 *M* = 1.9603, *SD* = .44872	-0.147 .456	

Notes: *Denotes ordinal score **Denotes significant correlation (α < .05) Correlations are between each multiple regression element and summative evaluation scores. Two relationships were significant: MTEBI and summative evaluation scores, and professional development and summative evaluation scores.

When professional development was introduced first, its *r* value was .327, with an r^2 value of .107, and an adjusted r^2 of .072, which was not significant (p = .090, α < .05). However, when math teaching efficacy was added, the new adjusted r^2 was .215 with that factor's individual significance at .025. The combined significance of the two was .019. Although professional development in isolation did not yield a significant relationship, the addition of mathematics self-efficacy did yield a significant relationship (α = .05), albeit at a lower level than with mathematics self-efficacy in isolation. The adjusted r^2 of .215 represented an effect size of .274, which is a medium-large effect size (Cohen, 1992; Durlak, 2009).

When the converse was analyzed, layering MTEBI first and then adding professional development second, the *r* value for MTEBI was .522 (p = .004, based on the *F* Change value). The r^2 value was .273, with an adjusted r^2 of .245. With professional development added, the r^2 change was .000. The initial model had a significance of .004 with just mathematics teaching efficacy, and when aggregated, the significance level was diminished to a value of .019 with professional development added. Professional development, therefore, had a mitigating impact on significance. The effect-size of this model was .325, which is considered large (Cohen, 1992; Durlak, 2009).

The aggregated regression formula was:

$$Summative\ score\ =\ -.545\ +\ .001*\ PD\ Level\ +\ .898\ *\ MTEBI$$

However, if only MTEBI was included, the formula was:

$$Summative\ score\ =\ -.550\ +\ .900\ *\ MTEBI$$

Inclusion of the professional development factor within the model appeared to be trivial, at best, even though the two-factor model reached an acceptable level of significance (α < .05). The

model's residuals were visually inspected for heteroscedasticity and did not show a problematic pattern. Collinearity was also analyzed (*VIF* = 1.635) and the level was not problematic.

In summary, the answers to question 3 supported by the study data are that mathematics self-efficacy has a significant relationship with summative evaluation scores, but that this relationship is mitigated by professional development. Additionally, mathematics professional development's relationship with summative evaluation scores becomes significant when teachers have higher levels of mathematics teaching self-efficacy.

Chapter V

Discussion

Introduction

Primary topics of this discussion include the study's sample demographics, survey approach, individual research questions, connections to the literature, limitations, and areas for further research. Because the Marzano system and VAM are relatively recent phenomena within the high-stakes accountability movement, this study would best be described as exploratory with regard to their effects within education. Much more research is needed pertaining to teacher evaluation. However, teacher self-efficacy, as a construct, is further developed. For example, to date, studies have not explored the triadic reciprocal relationship between mathematics self-efficacy, mathematics instructional professional development, and mathematics VAM. Because self-efficacy appears to play a pivotal role in this relationship, the study provides a new perspective.

Sample Demographics

Demographics of the sample were analyzed for the purpose of identifying possible generalizability to the district and state populations of fourth- and fifth-grade teachers. Elements investigated were gender, race, and years of teaching.

The sample's gender and racial demographics were significantly similar to those of the district from which the sample was drawn ($\alpha < .01$), but were different from those in the state of Florida. Given that the district's racial demographics are also significantly different from the state's ($\alpha < .01$), this result was to be expected. The sample's gender demographics were significantly similar to those of the district and the state ($\alpha < .05$). However, the district's gender

balance is significantly different from the state's ($\alpha < .01$), with a higher percentage of males in the district than in the state.

The sample's mean years of teaching experience, 15.25, exceeded both the district and state values of 12.13 years and 11.66 years, respectively. However, because state reporting (Florida Department of Education, 2014) provides only mean values for years of teaching, no further analysis of similarity was possible.

Due to the size of the sample, comparisons of the sample to the district and state may be limited. These possible limitations are discussed more fully in the limitations portion of this chapter.

Survey Approach

Related to the issue of sample size are the electronic survey methodology and sampling technique. For this study, the electronic survey link was sent to all fourth- and fifth-grade teachers within the sample school district. The approach yielded usable responses with links to evaluations for 32 teachers, which is only about a 27% response rate, even though reminders and a drawing incentive were used. Electronic surveys tend to yield higher response rates than paper surveys, and face-to-face interviews or observations would have caused either fewer subjects for the study due to lack of anonymity, or potential study bias due to a need for the researcher to interpret the interviews/observations. However, if a different methodology had been used, the sampling method could have been more strategic. For example, subjects with varied evaluation results could have been chosen for study at the outset. By choosing subjects strategically, the researcher might have mitigated the lack of Marzano score variance or evaluation score variance. Perhaps, future studies might use a mixed methods approach to better elicit this information.

Summary of Results

  Question 1. What is the relationship between teachers' levels of self-efficacy and teacher observation scores derived from the Marzano evaluation system? Question 1was explored through use of results from the TSES and MTEBI surveys. Neither set of survey data yielded significant correlations with Marzano observation results. Noteworthy in the data is a lack of variance in Marzano scores for the sample, with almost all subjects being rated as either Effective or Highly Effective. Perhaps a larger sample with more variance would yield different results. M. Toth, a researcher with the Marzano-affiliated Learning Sciences International (personal communication, January 19, 2015) has noted that the initial implementation of the Marzano evaluation model in Florida had a similar pattern, with lack of variance in scores. In one studied district, however, further observer training and assignment of administrators to teachers in other schools corresponded to increased fidelity within the scoring system. The tendency toward high scores, or ceiling effect, of the Marzano observation system's implementation is part of a larger picture. A recently published report by the Florida Department of Education (Florida Department of Education, 2014c) shows a similar trend with summative evaluation results for Florida teachers, with only 2.3% receiving ratings of Needs Improvement, Developing, or Unsatisfactory. Thus far, there has also been some relationship between Marzano scores and VAM in Florida (M. Toth, personal communication, January 19, 2015), but that the relationship has been stronger on short-term repeated student performance measures, such as unit tests. As the model becomes more grounded, this relationship may change. Given these state-wide circumstances, the lack of variance and relationship in this study was to be expected.

  The findings of this study do not necessarily mean that teacher self-efficacy and teaching performance are unrelated. In fact, Heneman, Kimball, and Milanowski (2007) did identify a

significant a relationship between self-efficacy levels and observation scores using the Danielson evaluation model. A larger sample with more variance in Marzano evaluation scores may yield a relationship. Additionally, the 2013-14 school year is the second year the subjects' school district used the Marzano model for observations. After the evaluation system has been implemented for a few more years and administrators have received further training, the scores may change. A later study may be warranted.

Noteworthy, also, is the fact that a significant correlation between Marzano observation scores and VAM was not identified. Although that correlation was not directly relevant to the questions of this research, a lack of relationship between observations and VAM is curious. A recent Florida study conducted with third-, fourth-, and fifth-grade teachers in a district near the subjects' county yielded similar findings regarding teacher observation scores and FCAT achievement (Flowers, 2013). In fact, the Flowers study identified significant, but modest, correlations between Marzano teacher observation scores and FCAT achievement in both reading and math for students of third-grade teachers, but found no such correlations for either fourth- or fifth-grade teachers. One other study found significant correlations between evaluation ratings and mathematics VAM scores for third- through eighth-grade teachers in Cincinnati, OH (Milanowski, 2004); this study, however, did not use the Marzano model. The divergent results between this study and other studies indicate that this is an area for further research. These results also point toward a reason for including both observation data and student performance data, because they appear to be measuring different teacher performance dynamics.

Question 2. What is the relationship between teachers' sense of self-efficacy and their Florida State value-added scores in mathematics? Question 2 was also explored through use of results from the TSES and MTEBI surveys. The dependent variable, however, was different; it

was the mathematics value-added score. There was no significant relationship between general teacher self-efficacy, as measured by the TSES, and VAM. However, a significant relationship was identified between mathematics teaching self-efficacy, as measured by the MTEBI, and mathematics VAM. One might expect that the two variables specific to mathematics would correlate, especially in light of Bandura's assertion (1997) that self-efficacy should be measured with specificity to the environment and task being studied. Self-efficacy is not static when tasks or circumstances change, so measurement in the mathematics-specific realm is appropriate.

Given the recent introduction of VAM for evaluation purposes, the relationship between mathematics self-efficacy and mathematics VAM is compelling. This relationship should be investigated further, perhaps with larger samples, to identify whether the relationship holds. Also interesting is that the mathematics outcome expectancy subscale correlated with mathematics VAM, although the personal mathematics teaching efficacy subscale did not. Some researchers suggest that the outcome expectancy subscale should not be included as part of the MTEBI instrument (Burns, 2009)**.** However, the finding that outcome expectancy and mathematics VAM correlate may cause one to reconsider whether abandoning the current MTEBI construct design would be prudent.

Question 3. What proportion of influence do the following have on a teacher's evaluation rating: teacher sense of efficacy, as measured by the Tschannen-Hoy instrument; mathematics teaching efficacy, as measured by the MTEBI; years of teaching experience; mathematics professional development; and mathematics coursework attained? As noted in Chapter IV, professional development and mathematics self-efficacy were explored using a hierarchical multiple regression to identify their relationships with summative evaluation scores for the teachers. Professional development did have a significant relationship with the evaluation

scores, and this relationship was strengthened by mathematics teaching self-efficacy. When the

hierarchical regression was run in the converse order, mathematics self-efficacy had a significant

relationship with evaluation scores, but this was mitigated by mathematics instruction

professional development. The phenomenon becomes even more interesting by the fact that that

efficacious teachers seem to participate in professional development with greater frequency than

non-efficacious teachers.

Several questions arise from these results. Why would mathematics self-efficacy correlate

with summative evaluations, while general self-efficacy does not? Mathematics teaching and

VAM compose a small portion of the summative evaluation, so one would expect general self-

efficacy to correlate, rather than the mathematics-specific measure. A second question is why

mathematics professional development might mitigate the relationship between mathematics

self-efficacy and evaluation results. Do efficacious teachers teach less well, in the eyes of their

supervisors, when they have had professional development? Perhaps they are trying new

techniques or taking more risks, which yield lower evaluation scores. Perhaps professional

development for teachers with strong efficacy should look different from professional

development conducted for low efficacy teachers. Each of these questions suggests areas for

future research.

Connections to Bandura's Model

Although the research questions did not specifically seek to identify whether the data

supported Bandura's triadic reciprocality model (Figure 1), such correlations became apparent as

the research was conducted. When identifying correlations for the questions, the researcher

noticed a pattern emerging that involved three particular significant correlations. Relationships

amongst mathematics teaching efficacy (MTEBI), mathematics professional development hours, and mathematics VAM were all significant.

Table 12

Correlations Between Mathematics Elements

	MTEBI	Math VAM
Math VAM	.527 (p = .006)	1.000
Math Professional Development	.602 (p = .000)[*]	.605 (p = .001) [*]

Note: [*] Correlations are Spearman rho *r* values

These correlations map quite readily to the Triadic Reciprocal Determinism construct of Bandura (1986). Within Bandura's model, there exist interrelationships between behavioral,

Figure 16

Connection Between Study Findings and Bandura's Model

Personal Factors – Biological, Cognitive and Affective
MTEBI results

ρ = .602, p = .000

r = .527, p = .006

ρ = .605, p = .001

Behavioral Factors –
What we do
   Participation in mathematics
   instruction professional
   development

Environmental Factors –
What occurs around us
Student achievement in
   mathematics as measured
   by VAM
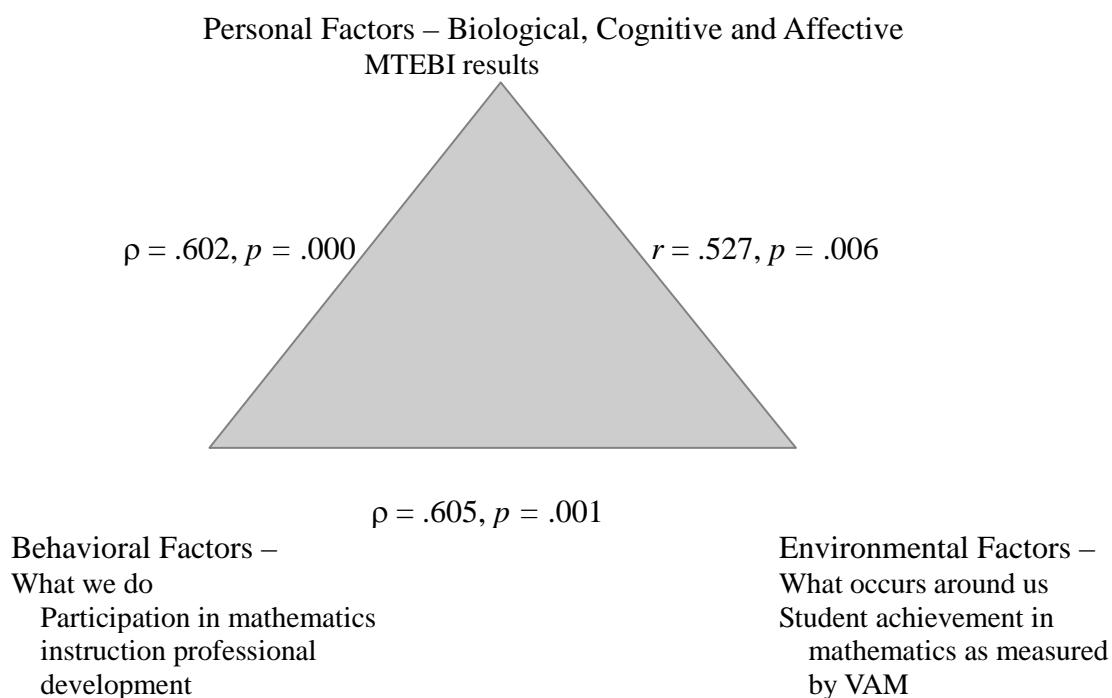
Note. Three study elements correlate with one another supporting Bandura's triadic reciprocal framework for mathematics teaching (1986, p. 24).

environmental, and personal factors. One study by Williams and Williams (2010) does verify the reciprocal determinism theory in a study with a large sample size that takes place in 30 countries. The approach isolated each variable to identify the direction of influence of each variable on the other factors in Bandura's model; as a result, they were able to suggest a causal effect of behavior on beliefs, as well as the converse. The international nature of this study also showed the triadic reciprocal determination model is not necessarily prone to cultural bias (Williams & Williams, 2010). A recent study (Muijs & Reynolds, 2015) also explored elements of reciprocal determinism; however, this study included a number of additional variables related to teaching style and pre- and post- tests of pupils.

For this study, the components that showed significant relationships were professional development participation (a behavioral factor), student performance (an environmental factor), and mathematics self-efficacy beliefs (a personal factor). Although the relationships found within this study support Bandura's model, this finding was incidental, and not related specifically to the research questions. The relationships exist and they support the model, nonetheless (see Figure 16). This connection between Bandura's model and three of the elements of the study, although not surprising, is worthy of future exploration.

Limitations

Sample size. The size of the sample is sufficient for correlational analysis (Salkind, 2012) for questions 1 and 2. Additionally, because only two factors were significant in the multiple regression (question 3), the sample size is also sufficient (Salkind, 2012). Had there been three or more significant contributors for the multiple regression, a larger sample would have been necessary. Although a large sample is always preferred, the sample size was acceptable.

Had a larger sample been available, the racial demographics of the sample may have been more similar to the state teacher population. Had the sample become more similar to the state numbers, however, the sample may have become dissimilar from the district racial composition. Although race and gender demographics are factors to control for, the literature indicates that teacher self-efficacy beliefs do not typically vary depending on race and gender (Tschannen-Moran & Hoy, 2007). Therefore, sample size does not necessarily pose a threat to generalizability on the basis of race or gender.

The pool of fourth- and fifth-grade teachers in the sample district represented approximately 110 individuals. Forty-six survey respondents accounted for about a third of the possible subjects. A number of subjects, however, did not complete the survey, yielding a usable sample of 32 teachers. Thirty-two teachers represented about a quarter of the possible subjects, which is not unusually low for a survey. Respondents who did not complete the survey ceased to participate at either the consent screen or the screen where the code needed to be entered to link the survey to evaluation results. One may presume, therefore, that some subjects may have stopped participating because they did not want to release evaluation results to the research. Adding more subjects would have enhanced the study. Although the number of subjects was not insufficient for a correlation study, *t* tests were introduced as a corollary, and the sample size was small for an inferential approach. Future studies of this type should involve larger samples, including predominantly urban and rural school systems.

Another limitation, now inherent in self-efficacy research is the prevalence of self-efficacy research. Subjects for the study are likely familiar with the construct of self-efficacy and survey results may be influenced by teacher bias, introduced by professional development or training regarding self-efficacy. Subjects may have believed there were *right answers* to the

questions. By introducing the study topic as teacher beliefs, rather than self-efficacy, an attempt was made to control for this bias. However, Kleinsasser (2014) reminds us that this issue exists with current teacher self-efficacy research, nonetheless.

Strategy. Other possible approaches for further sampling and enhanced generalizability may include strategic sampling or interviews. Bandura (2006) supports the self-reporting approach to gathering self-efficacy information but suggests that there are inherent risks of inaccuracy with self-reporting that can be mitigated with use of other approaches. Interviews may yield more accurate information or may add to the richness of a study, and the use of observations might mitigate risks of self-reporting inaccuracies. However, interviews and observations would add researcher bias, which is not standard for a quantitative study (Marshall & Rossman, 2011).

Recommendations for Further Research

Although this study was exploratory, a number of key findings were identified. There was a significant correlation between mathematics instructional self-efficacy and mathematics VAM. Mathematics instruction professional development was also found to have a significant relationship with both mathematics VAM and mathematics instructional self-efficacy. Mathematics self-efficacy's relationship with summative evaluation scores was mitigated by mathematics instruction professional development. Conversely, mathematics instruction professional development's relationship with summative evaluation scores was found to be enhanced by mathematics teaching self-efficacy. These relationships support Bandura's triadic reciprocal model (1986). Although the sample size for this study was sufficient to reach these findings, future studies with larger samples and more variance in Marzano and summative evaluation scores should be conducted. Sampling methods should include subjects that more

closely mirror the population of the state or nation; this approach would increase the ability to generalize the findings.

This study adds to the body of literature regarding value-added data and the Marzano observation system. However, further research must be conducted involving these approaches, partially because of their recent implementation, but also due to the high-stakes decisions being tied to these systems. This study points out some dilemmas within value-added teacher evaluation systems, namely the lack of relationship between Marzano scores and any other metric, as well as the lack of relationship between the Marzano system and VAM.

Mixed methods studies, in the future, may yield further information about the triadic reciprocal theory as applied to mathematics instruction. By interviewing or observing subjects, information may be gleaned regarding how self-efficacy, actions, and environment interact. This information may help develop a theory of action for instructors and administrators as they attempt to improve student achievement in mathematics.

Implications for Practitioners

This study is best described as exploratory. While the findings are significant, the sample size and correlation approach are limitations. Additionally, value-added modeling and Marzano observation results as dependent variables are also not fully vetted within the literature. Due to these limitations and the novelty of the variables, any recommendation for practitioners or professional developers is offered with appropriate caution (Gersten, 2001; Stanovich & Stanovich, 2003).

However, the findings suggest several intriguing possibilities. Content-specific teacher efficacy beliefs do appear to have a relationship with value-added scores, and overall summative evaluation scores. Efficacy levels also appear to be the primary influencer on mathematics VAM

when combined with professional development. As a result, professional developers, school based administrators, and curriculum specialists should at least consider the efficacy levels of teachers when planning staff performance improvement initiatives. Bandura (1977, 1986) identifies four influences on efficacy levels: mastery experiences, vicarious experiences, verbal persuasion, and physiological factors. *Job embedded professional development* (National Comprehensive Center for Teacher Quality, 2010)*,* has potential. Tenets of job-embedded professional development include lesson review, ongoing dialogue about practice, coaching, and collegial dialogue. These practices allow for mastery experiences, vicarious experience, verbal persuasion and positive physiological affective arousal.

Pre-service experiences that involve actual teaching experiences early and often during reflective coursework may also have potential. Earlier teaching experiences may *front load* mastery experiences and help develop appropriate stress responses to teaching. When practice teaching is combined with reflective dialogue with fellow students, coaches, or professors teaching self-efficacy levels can be impacted.

Further research should be conducted regarding efficacy levels and job-embedded professional development as well as embedded pre-service models to verify whether these theoretical ideas show success. Efficacy levels for pre-service teachers do show instability (Tschannen-Moran & Hoy, 2007); perhaps earlier teaching experiences within the pre-service program would help establish stability of efficacy levels earlier in teachers' careers.

Epilogue

Maria has participated in district sponsored and school sponsored mathematics professional development with her grade-level team at Grapefruit Grove Elementary. This professional development has boosted her confidence in teaching mathematics and has given her

some new strategies to try with her students. Some of these new approaches have been well received by her students, and they have performed well on formative assessments. The successes have been shared with her grade-level team, and there is an evolving synergy around improved mathematics instruction. Maria's principal has taken notice, complimenting her in her last few observations. Professional development, classroom practice, Maria's self-efficacy, and mathematics achievement all seem interrelated. Where one is lacking, they all seem to be lacking, but when there are successes, they all seem to improve. The benefits of this improvement are Maria's students, who are better able to learn and enjoy mathematics.

References

Abbott, M., Baker, D., Smith, K., & Trzyna, T. (2010). *Winning the math wars: No teacher left behind.* Seattle, WA: University of Washington Press.

Aerni, P. W. (2008). *Teacher self-efficacy and beliefs for teaching mathematics in inclusive settings* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 304446066)

Althauser, K. L. (2010). *The effects of a sustained, job-embedded professional development on elementary teachers' math teaching self-efficacy and the resulting effects on their students' achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. ED529661)

American Statistical Association. (2014, April 18). *ASA statement on using value-added models for educational assessment.* Retrieved from http://www.amstat.org/policy/pdfs/ASA_VAM_statement.pdf

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65-75.

Anderson, R. N., Greene, M. L., & Loewen, P. S. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy and student achievement. *The Alberta Journal of Educational Research, 34*(2), 148-165.

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnel, L., Pascal, A., … Zellman, G. L. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools* (No. R-2007-LAUSD). Santa Monica, CA: RAND.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. White Plains, NY: Longman.

Bakar, A. R., Mohamed, S. S., & Zakaria, N. S. (2012). They are trained to teach, but how confident are they? A study of student teachers' sense of efficacy. *Journal of Social Sciences, 8*(4), 497-504. doi:10.3844/jsssp.2012.497.504

Baker, B. D., Oluwole, J. O., & Green III, P. C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education policy analysis archives*, *21*(5), 1-71.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator, 29*(3) Fall, 14-46.

Baeder, J. (2011, June 20). Marzano's "causal" evaluation system [Web log post]. Retrieved from http://blogs.edweek.org/edweek/on_performance/2011/06/marzanos_causal_ evaluation_system.html

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, *84*, 191-215. doi:10.1037//0033-295X.84.2.191

Bandura, A. (Ed.). (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), *Annals of child development: Volume 6* (pp. 1-60). Greenwich, CT: JAI Press.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*, 248-287. doi:10.1016/0749-5978(91)90022-L

Bandura, A. (Ed.). (1995). *Self-efficacy in changing societies*. New York, NY: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W.H. Freeman and Company.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, *9*(3), 75-78. doi:10.1111/1467-8721.00064

Bandura, A. (2004). Swimming against the mainstream: The early years from chilly tributary to transformative mainstream. *Behaviour Research and Therapy, 42*. 613-630. doi:10.1016/j.brat.2004.02.001

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F, Pajares & T. C. Urdan (Eds.), *Self-efficacy beliefs of adolescents*, *5*(pp. 307-337). Phoenix, AZ: Information Age Publishing, Inc.

Bartz, A. (1976). *Basic statistical concepts in education and in the behavioral sciences*. Minneapolis, MN: Burgess.

Berman, P., McLaughlin, M., Bass, G., Pauly, E., & Zellman, G. (1977). *Federal programs supporting educational change: Vol. VII. Factors affecting implementation and continuation* (No. R-1589/7-HEW). Santa Monica, CA: RAND.

Bong, M., & Clark, R. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist, 34*(3), 139-153. doi:10.1207/s15326985ep3403_1

Boonen, T., Pinxten, M., Van Damme, J., & Onghena, P. (2014). Should schools be optimistic? An investigation of the association between academic optimism of schools and student achievement in primary education. *Educational Research & Evaluation*, *20*(1), 3-24. doi:10.1080/13803611.2013.860037

Braun, H. (2013). Value-added modeling and the power of magical thinking. *Ensaio: Avaliação e Políticas Públicas em Educação*, *21*(78), 115-130. doi:10.1590/s0104-40362013000100007

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Retrieved from Educational Testing Service (ETS) website: www.ets.org/research/pic

Bruce, C. (2005, October). *Teacher candidate efficacy in mathematics: Factors that facilitate increased efficacy.* Paper presented at 27th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Blacksburg, VA.

Bruce, C. D., & Ross, J. A. (2008). A model for increasing reform implementation and teacher efficacy: Teacher peer coaching in grades 3 and 6 mathematics. *Canadian Journal of Education, 31*(2), 346-370.

Byrne, B. M. (1984). The general/academic self-concept/nomological network: A review of construct validation research. *Review of Educational Research, 54*(3), 427-456. doi:10.3102/00346543054003427

Carson, P. B. (2012). *Effects of levels of formal educational training in mathematics on teacher self-efficacy.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 1035351641)

Cetinkaya, B., & Erbas, A. K. (2011). Psychometric properties of the Turkish adaptation of the Mathematics Teacher Efficacy Belief Instrument for in-service teachers. *The Spanish Journal of Psychology, 14*(2), 956-66. doi:10.5209/rev_SJOP.2011.v14.n2.41

Cohen, J. (1988). *Statistical Power Analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1). 155-159. doi:10.1037/0033-2909.112.1.155

Cook v. Stewart, No. 1:13-cv-72 (N.D. Fla. 2014).

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed). Boston, MA: Pearson Education.

Copa, J. (2012, November) *Florida's value-added model: Overview of the model used to measure student learning growth on FCAT reading and mathematics.* Presentation at Florida Educational Research Association (FERA) Annual Meeting, Gainesville, FL. Retrieved from

http://www.feraonline.org/presentations/FloridaValueAddedModelJuanCopaFLDOE.pdf

Davies, B. (2004). *The relationship between teacher efficacy and higher order instructional emphasis* (Conference Proceeding No. DAV04854). Melbourne, AU: NSW Department of Education and Training.

DeMitchell, T. A., DeMitchell, T. A., & Gagnon, D. (2012). Teacher effectiveness and value-added modeling: Building a pathway to educational malpractice. *Brigham Young University Education and Law Journal*, 257-301.

Denzine, G. M., Cooney, J. B., & McKenzie, R. (2005). Confirmatory factor analysis of the teacher efficacy scale for prospective teachers. *British Journal of Educational Psychology, 75*, 689-708. doi:10.1348/000709905X37253

Dively Jr, J. A. (2012). Growth model evaluations: Possibilities and pitfalls. *Faculty Research and Creative Activity.* Paper 1. Retrieved from http://thekeep.eiu.edu/edadmin_fac/1

Doherty, K. M., & Jacobs, S. (2013). State of the states 2013: *Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. Report retrieved from the National Council on Teacher Quality website: http://www.nctq.org/dmsView/State_of_the_States_2013_Using_ Teacher_Evaluations_NCTQ_Report

Domsch, G. D. (2009). *A study investigating relationships between elementary principals' and teachers' self-efficacy and student achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Thesis Database. (UMI: 3383305)

Drew, C. J., Hardman, M. L., & Hosp, J. L. (2008). *Designing and conducting research in education*. Los AngelesCA: Sage.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology*, *34*(9). 917-928. Retrieved from http://jpepsy.oxfordjournals.org/content/34/9/917.full.pdf+html

Eells, R. J. (2011). *Meta-analysis of the relationship between collective teacher efficacy and student achievement*. (Doctoral dissertation). Retrieved from http://ecommons.luc.edu/luc_diss/133

Elliott, E. M., Isaacs, M. L., & Chugani, C. D. (2010). Promoting self-efficacy in early career teachers: A principal's guide for differentiated mentoring and supervision. *Florida Journal of Educational Administration & Policy*, *4*(1), 131-146. Retrieved from http://www.eric.ed.gov/PDFS/EJ911436.pdf

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics, 100*(4), 194-202. doi:10.1111/j.1949-8594.2000.tb17256.x

Esterly, E. J. (2003). *A multi-method exploration of the mathematics teaching efficacy and epistemological beliefs of elementary preservice and novice teachers* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 305320293)

Farmani, S. & Khamesan , A. (2011). Validating Mathematics Teaching Efficacy Beliefs Instrument among pre-service teachers in primary schools. *Innovation in Management Education (Journal of Modern Thoughts in Education), 6*(4). 22-31.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Thousand Oaks, CA: Sage.

Fives, H., & Buehl, M. M. (2010). Examining the factor structure of the Teachers' Sense of Efficacy Scale. *The Journal of Experimental Education, 78*(1), 118-134. doi:10.1080/00220970903224461

Florida Department of Education. (2011). *Data report: Free/reduced price lunch eligibility.* (Education Information and Accountability Services Series 2011-03D). Retrieved from http://www.fldoe.org/eias/eiaspubs/pdf/frplunch0910.pdf

Florida Department of Education. (2014a). *Data publications and reports: Students.* Retrieved from http://www.fldoe.org/eias/eiaspubs/pubstudent.asp

Florida Department of Education. (2014b). *Data publications and reports: Teachers.* Retrieved from http://www.fldoe.org/eias/eiaspubs/word/tchdemo1314.doc

Florida Department of Education. (2014c). *2013-14 Preliminary district educator evaluation results 11-19-14*. Retrieved from http://www.fldoe.org/core/fileparse.php/5423/urlt/1314EduEvalRatings.pdf

Florida Department of Education, American Institutes for Research. (2012). *Florida value-added technical report* [Working draft]. Retrieved from www.fldoe.org/committees/.../value-added-model-technical-report.docx

Flowers, A. R. (2013). *A study of the Marzano teacher evaluation model and student achievement at 24 elementary schools in a large suburban school district in Central Florida* (Doctoral dissertation). Retrieved from http://etd.fcla.edu/CF/CFE0004841/FLOWERS_AMY_R201308_EDD-FINAL.pdf

German, J. M. (2014). *Teachers' perceptions of self-efficacy: The impact of teacher value-added* (Doctoral dissertation). Retrieved from https://etd.ohiolink.edu/rws_etd/document/get/ashland1398439686/inline

Gersten, R. (2001). Sorting out the roles of research in the improvement of practice. *Learning Disabilities Research & Practice*, *16*(1), 45-50.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 75*(4), 569-582. doi:10.1037//0022-0663.76.4.569

Gist, M., & Mitchell, T. R. (1992, April). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review, 17*(2), 183. doi:10.2307/258770

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, *37*(2), 479-507.

Goldring, E., Grissom, J. A., Neumerski, C., Cannata, M., Rubin, M., Drake, T., & Schuermann, P. (2014, March). *Moving on out: Value-added measures: Principals' human capital decisions and the emergence of teacher observation data*. Paper presented at the annual meeting of the Association for Education Finance and Policy, San Antonio, TX.

Grant, R. W., & Sugarman, J. (2004). Ethics in human subjects research: Do incentives matter? *Journal of Medicine and Philosophy*, *29*(6), 717-738. doi:0360-5310/04/2906-717

Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31*(3), 627-643. doi:10.3102/00028312031003627

Hansford, B., & Hattie, J. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52,* 123-142. doi:10.3102/00346543052001123

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267-71.

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, *51*(1), 73-112.

Haystead, M. W., & Marzano, R. J. (2009). *Meta-analytical synthesis of studies conducted at Marzano Research Laboratory on instructional strategies*. Englewood, CO: Marzano Research Laboratory.

Head, C. F. (2012). *Comparative analyses of mathematics teachers' efficacy using factor analysis and the Rasch Model*. (Doctoral dissertation). Retrieved from Dissertations, Theses and Capstone Projects Database. (Paper No. 526)

Heneman, H. G., III, Kimball, S., & Milanowski, A. (2006). *The Teacher Sense of Efficacy Scale: Validation evidence and behavioral prediction* (WCER Working Paper No. 2006-7). Retrieved from http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2006_07.pdf

Henson, R. K. (2001, January). *Teacher self-efficacy: Substantive implications and measurement dilemmas*. Invited keynote address given at the Annual Meetings of the Educational Research Exchange, Texas A & M. College Station, TX.

Henson, R. K., Kogan, L. R., & Vacha-Hasse, T. (2001). A reliability generalization study of the teacher efficacy scale and related instruments. *Educational and Psychological Measurement, 61*(3), 404. doi:10.1177/00131640121971284

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education, 41*(5), 513-545.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430-511.

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371-384.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*(3), 794-831. doi:10.3102/0002831210387916

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology*, *105*(3), 774-786. doi:10.1037/a0032198

Howe, K. R., & Moses, M. S. (1999). Ethics in educational research. *Review of Research in Education*, *24*, 21-59. Retrieved from http://www.colorado.edu/education/ sites/default/files/attached-files/Howe_Moses_Ethics_in_Educational_Research.pdf

Hoy, A. W. (2000, April). *Changes in teacher efficacy during the early years of teaching*. Paper Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, *43*(3), 425-446.

Isiksal, M. (2005). Pre-service teachers' performance in their university coursework and mathematical self-efficacy beliefs: What is the role of gender and year in program? *Mathematics Educator, 15*(2), 8-16.

Isiksal, M. (2010). The relationship among mathematics teaching efficacy, math anxiety, and mathematical self-concept: The case of Turkish pre-service elementary teachers. *Asia-Pacific Education Researcher, 19*(3), 501-514. doi:10.3860/taper.v19i3.1857

Isiksal, M., & Cakiroglu, E. (2005). Teacher efficacy and academic performance. *Academic Exchange Quarterly,* (Winter), 28.

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 398-407.

Jones, D. L. (2012). *The effects of content knowledge on teacher efficacy of secondary mathematics teachers within three mid-south states* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 945732837)

Kanold, T. D., & Larson, M. R. (2012). *Common Core Mathematics in a PLC at Work: Leader's Guide*. Bloomington, IN: Solution Tree Press.

Kieftenbeld, V., Natesan, P., & Eddy, C. (2010). An item response theory analysis of the Mathematics Teaching Efficacy Beliefs Instrument. *Journal of Psychoeducational Assessment*, *29*(5). 443-454.

Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology*, *34*(1), 67-76.

Klassen, R. M., Tze, V. M . C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998-2009: Signs of progress or unfulfilled promise? *Educational Psychology Review,* *23*(1), 21-43. doi:10.1007/s10648-010-9141-8

Kleinsasser, R. C. (2014). Teacher efficacy in Teaching and Teacher Education. *Teaching and Teacher Education*, *44*, 168-179. doi:10.1016/j.tate.2014.07.007

Koedel, C. & Betts, J. R. (2007). *Re-examining the role of teacher quality in the education production function.* (Working Paper No. 2007-708). Retrieved from Missouri Department of Economics website: http://economics.missouri.edu/working-papers/2007/wp0708_koedel.pdf

Laerd Statistics. (2014). *IBM Tutorials and Statistical Guides.* Retrieved from https://statistics.laerd.com/

Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association*,*73*(362), 253-263.

Larson, M. R. (2012). Will CCSSM have staying power? *Journal of mathematics education at teacher's college 3*(1). 50-53. Retrieved from http://journals.tc-library.org/index.php/matheducation/article/viewFile/795/490

Maguire, K. (2011). *The role of teacher efficacy in student academic achievement in mathematics* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 863675545)

Marsh, H. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*(4), 623-636. doi:10.1037/0022-0663.82.4.623

Marsh, H., Walker, R., & Debus, R. (1990, September). *Subject specific components of academic self concept and self-efficacy: A response to Skaalvik and Rankin.* Paper presented at the Australian Association for Research in Education Conference, Sydney, Australia. Retrieved from http://www.aare.edu.au/data/publications/1990/walke90s244.pdf

Marshall, C., & Rossman, G. B. (2011). *Designing qualitative research.* Thousand Oaks, CA: Sage.

Marzano, R. J. (2007). *Art and science of teaching: A comprehensive framework for effective instruction.* Alexandria, VA: Association for Supervision & Curriculum Development (ASCD).

Marzano, R. J. (2011). Marzano Art and Science of Teaching Teacher Evaluation Model. Retrieved from http://www.marzanoevaluation.com/files/Day_1/LearningMap_4Domains.pdf

Marzano, R. J. (2013). *The Marzano teacher evaluation model.* Englewood, CO: Marzano Research Laboratory. Retrieved from www.marzanoresearch.com

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching.* Alexandria, VA: Association for Supervision & Curriculum Development (ASCD).

Marzano, R. J., & Toth, M. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. Alexandria, VA: Association for Supervision & Curriculum Development (ASCD).

Marzano, R. J., Toth, M., & Schooling, P. (2012). *Examining the role of teacher evaluation in student achievement: Contemporary research base for the Marzano causal teacher evaluation model.* West Palm Beach, FL: Learning Sciences International. Retrieved from Marzanocenter.com

McCaffrey, D., Lockwood, J. R. Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND. Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG158.pdf

McCoy, A. C. (2011). *Specialized mathematical content knowledge of preservice elementary teachers: The effect of mathematics teacher efficacy* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 881786139)

Mertler, C. A., & Charles, C. M. (2011). *Introduction to Educational Research* (7th ed.). Boston, MA: Pearson.

Mewborn, D. (2013). Teachers content knowledge, teacher education, and their effects on the preparation of elementary teachers in the United States. *Mathematics Teacher Education and Development (MTED)*, *3*. 28-36. Retrieved from http://www.merga.net.au/ojs/index.php/mted/article/viewFile/120/96

Midgley, C., Feldlaufer, H., & Eccles, J. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81*(2), 247-258. doi:10.1037/0022-0663.81.2.247

Mills, D. E. (2009). *A study of the relationship between collective teacher efficacy and the elementary mathematics achievement gap* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 250178823)

Moore, M., & Esselman, M. (1992, April). *Teacher efficacy, empowerment, and a focused instructional climate: Does student achievement benefit?* Paper presented at the Annual Conference of the American Educational Research Association, San Francisco, CA.

Muijs, D., & Reynolds, D. (2015). Teachers' beliefs and behaviors: What really matters? *Journal of Classroom Interaction, 50*(1), 25-40.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38,* 30-38. doi:10.1037/0022-0167.38.1.30

National Comprehensive Center for Teacher Quality. (2010, April). *Job-embedded professional development: What it is, who is responsible, and how to get it done well.* (Issue Brief). . Washington, D.C.: Croft, A., Coggshall, J. G., Dolan, M., & Powers, E. Retrieved from http://files.eric.ed.gov/fulltext/ED520830.pdf

Newton, K. J., Leonard, J., Evans, B. R., & Eastburn, J. A. (2012). Preservice elementary teachers' mathematics content knowledge and teacher efficacy. *School Science and Mathematics*, *112*(5), 289-299. doi:10.1111/j.1949-8594.2012.00145.x

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), 1-27. doi:http://dx.doi.org/10.14507/epaa.v18n23.2010

Nisbet, S., & Warren, E. (2013). Primary school teachers' beliefs relating to mathematics, teaching and assessing mathematics and factors that influence these beliefs. *Mathematics Teacher Education and Development*, *2,* 34-47. Retrieved from http://www.merga.net.au/ojs/index.php/mted/article/viewFile/128/104

O'Keeffe, J. (2012, December 1). Marzano is God [Web log post]. Retrieved from http://www.dissidentteacher.com/2012/12/marzano-is-god.html

Phelps, C. M. (2009). *Investigating the development of pre-service elementary teachers' mathematics self-efficacy beliefs and learning goals: A review of research and a mixed methods study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (Accession No: ED512774)

Postal, L. (2012, July 13). Teachers: New evaluation system 'artificial,' 'frustrating,' 'humiliating'. *The Orlando Sentinel.* Retrieved from http://articles.orlandosentinel.com/2012-07-13/features/os-florida-teacher-evaluations-20120713_1_new-teacher-evaluations-evaluation-plan-teacher-merit-pay-law

Provasnik, S., Gonzales, P., & Miller, D. (2009). *U.S. performance across international assessments of student achievement: Special supplement to the condition of education 2009* (NCES 2009-083). Retrieved from the National Center for Education Statistics website: nces.ed.gov.

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of U.S. fourth-and eighth-grade students in an international context.* (NCES 2013-009). Retrieved from the National Center for Education Statistics website: nces.ed.gov.

Qualtrics. (2014). [Computer software] Provo, UT: Qualtrics, LLC.

Reilly, E., Dhingra, K., & Boduszek, D. (2014). Teachers' self-efficacy beliefs, self-esteem, and job stress as determinants of job satisfaction. *International Journal of Educational Management*, *28*(4), 365-378.

Ross, J. A. (1995). Strategies for enhancing teachers' beliefs in their effectiveness: Research on a school improvement hypothesis. *Teachers College Record, 97*(2), 227-250.

Rothstein, R., Ladd, H. F., Ravitch, D., Baker, E. L., Barton, P. E., Darling-Hammond, L, …Shepard, L. A. (2010, August 27). *Problems with the use of test scores to evaluate teachers* (Briefing Paper #278). Retrieved from the Economic Policy Institute website: http://www.epi.org/publication/bp278/

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*.(1), 1. doi:10.1037/h0092976

Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, *8*(3), 255-281.

Salkind, N. (2013). *Statistics for people who (think they) hate statistics* (5th ed.). Thousand Oaks, CA: Sage.

Shaughnessy, M. F. (2004). An interview with Anita Woolfolk: The educational psychology of teacher efficacy. *Educational Psychology Review, 16*(2), 153-176. doi:10.1023/B:EDPR.0000026711.15152.1f

Singer, E., & Bossarte, R. M. (2006). Incentives for survey participation: When are they "coercive"? *American Journal of Preventive Medicine*, *31*(5), 411-418. doi:10.1016/j.amepre.2006.07.013

Skaalvik, E. M. & Skaalvik, S. (2014). Teacher self-efficacy and perceived autonomy: Relations

    with teacher engagement, job satisfaction, and emotional exhaustion. *Psychological*

    Reports: Employment Psychology & Marketing, 114(1). 68-77. doi:

    10.2466/14.02.PR.114k14w0

Stanovich, P. J., & Stanovich, K. E. (2003). *Using research and reason in education: How*

    teachers can use scientifically based research to make curricular & instructional

    decisions. Portsmouth, NH: RMC Research Corporation. Retrieved from

    http://files.eric.ed.gov/fulltext/ED482973.pdf

Stewart, B. E. (2006). *Value-added modeling: The challenge of measuring educational outcomes.*

    Retrieved from Carnegie Corporation of New York website:

    http://carnegie.org/fileadmin/Media/centennialmoments/value_added_chal_paper_mocku

    p.pdf

Stigler, J., & Hiebert, J. (1997, October). *Understanding and improving classroom instruction:*

    An overview of the TIMSS video study. Paper Presented at the Australian Council for

    Educational Research Conference, Melbourne, Australia. Retrieved from

    http://www.acer.edu.au/documents/TIMSS_97-InsightsFromTIMSS.pdf#page=59

Stigler, J. W., & Hiebert, J. (2009). *The teaching gap: Best ideas from the world's teachers for*

    improving education in the classroom. New York, NY: Simon and Schuster.

Student Success Act., Fla. Stat. Ann. § 1012.34(3) (2011).

Swackhamer, L. E., Koellner, K., Basile, C., & Kimbrough, D. (2009). Increasing the self-

    efficacy of inservice teachers through content knowledge. *Teacher Education Quarterly*,

    36(2), 63-78. Retrieved from http://www.eric.ed.gov/PDFS/EJ857476.pdf

Swan, B. G., Wolf, K. J., & Cano, J. (2011). Changes in teacher self-efficacy from the student teaching experience through the third year of teaching. *Journal of Agricultural Education, 52*(2), 128-139. doi:10.5032/jae.2011.02128

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53-55.

Thames, M. H., & Ball, D. L. (2013). Making Progress in US Mathematics Education: Lessons Learned—Past, Present, and Future. In K. R. Leatham (Ed.), *Vital directions for mathematics education research* (pp. 15-44). New York: Springer doi:10.3102/00346543068002202

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 83-805. doi:10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and teacher Education*, *23*(6), 944-956.

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *The Elementary School Journal, 110*(2), 228-245. doi:10.1086/605771

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248.

Uitenbroek, D. G. (1997). *SISA Binomial.* Hilversum, Netherlands: D. G. Uitenbroek. Retrieved from http://www.quantitativeskills.com/sisa/statistics/fiveby2.htm

U.S. Census Bureau. *United States Census 2010*. Retrieved from https://www.census.gov/2010census/popmap/ipmtext.php?fl=12

U.S. Department of Education. (2009). *Race to the Top program executive summary.* Retrieved

from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Health & Human Services, National Commission for the Protection of

Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report:*

Ethical principles and guidelines for the protection of human subjects of research.

Washington, DC: US Department of Health, Education, and Welfare. Retrieved from

http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Vieluf, S., Kunter, M., & van de Vijver, F. R. (2013). Teacher self-efficacy in cross-national

perspective. *Teaching & Teacher Education*, *35*. 92-103. doi:10.1016/j.tate.2013.05.006

Wainer, H. (2011). Value-added models to evaluate teachers: A cry for help. *Chance, 24*(1), 11-

13. doi:10.1007/s00144-011-0003-z

Ward, E. K., & Johnston, E. (2014). A framework for revising the mathematics teaching efficacy

beliefs instrument. In Matney, G. T. & Che, S. M. (Eds), *Proceedings for the 41st Annual*

Meeting of the Research Council on Mathematics Learning, San Antonio, TX, (pp.64-71).

Retrieved from http://web.unlv.edu/RCML/RCML%20Proceedings(2014)%20Final.pdf

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of*

research tells us about the effect of leadership on student achievement. Aurora, CO: Mid-

continent Research for Education and Learning.

Williams, D. M. (2010). Outcome expectancy and self-efficacy: Theoretical implications of an

unresolved contradiction. *Personality and Social Psychology Review, 14,* 417.

doi:10.1177/1088868310368802

Williams, T., & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology, 102*(2), 453-466. doi:10.1037/a0017271

Wilson, D. M. (2013, Winter). Do the math! *Harvard Ed. Magazine.* Retrieved from http://www.gse.harvard.edu/ed/winter-2013/articles.

Woolfolk-Hoy, A. (2000, April). *Changes in teacher efficacy during the early years of teaching.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, L.A. Retrieved from http://anitawoolfolkhoy.com/pdfs/efficacy-43-22.pdf

Woolfolk, A. E., & Hoy, W. K. (2001) *Teacher efficacy scale: Short form.* Retrieved from http://people.ehe.osu.edu/ahoy/files/2009/02/tses.pdf

Woolfolk, A. E., & Hoy, W. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*(1), 81-91. doi:10.1037/0022-0663.82.1.81

Wyatt, M. (2012). Towards a re-conceptualization of teachers' self-efficacy beliefs: tackling enduring problems with the quantitative research and moving on. *International Journal of Research and Method in Education, 35*(2), 1-24. doi:10.1080/1743727X.2012.742050

Zambo, R., & Zambo, D. (2008). The impact of professional development in mathematics on teachers' individual and collective efficacy: The stigma of underperforming. *Teacher Education Quarterly, 35*(1), 159-168.

Appendix A

Demographic Analysis

Staff Race Demographic Percentages (Florida Department of Education, 2014b)

	Sample County	State of Florida	Sample
White Male	16.8%	12.2%	12.5%
Black Male	2%	2.5%	0%
Hispanic Male	1.6%	2.0%	0%
White Female	70.5%	46.7%	81.25%
Black Female	4.4%	8.5%	3.13%
Hispanic Female	3.7%	9.0%	3.13%

Based on Fisher's exact test, $(2, N = 32)$ p values for the racial comparison were the following:

> Sample compared with state $p = .00352$ (not significantly similar $\alpha < .01$)
> Sample compared with school district $p = .49085$ (significantly similar $\alpha < .05$)

For the district racial comparison with the state, a chi-square goodness of fit test was used due to the larger values.

> District compared with state chi-square value $(2, N = 32) = 127.771$, $p = .000$
> (not significantly similar, $\alpha < .01$)

For each gender comparison, a chi-square goodness of fit test was used $(1, N = 32)$.

> Sample compared with state chi-square $= .475$ $p = .49064$ (significantly similar $\alpha < .05$)
> Sample compared with district chi-square value $= 1.261$, $p = .26154$
> (significantly similar $\alpha < .05$)
> District Compared with state chi-square value $= 9.395$, $p = .00218$
> (not significantly similar, $\alpha < .01$)

Appendix B

Permission to Use Triadic Reciprocal Determinism Diagram (Figure 1)

From: wfritz@nnu.edu William Fritz
To: albertob@stanford.edu
Date: Tue, 31 Dec 2013 08:59:56 -0500
Subject: Permission Request to Use Triadic Reciprocality Chart

Dear Dr. Bandura:

I am a Doctoral student at Northwest Nazarene University (Nampa, ID), I am
in the initial stages of writing my dissertation. My topic of investigation
is mathematics teaching-efficacy beliefs of 4th grade teachers, and
relationships between these beliefs and classroom performance using both
the Marzano teacher evaluation system as well as State of Florida Value
Added Model scores.

I am in the initial planning stages now, and would like to request your
permission to utilize your Diagram (slightly modified) demonstrating
Triadic Reciprocality Construct (Bandura 1986) within my Literature Review
to add clarity to my explanation of self-efficacy and Social Cognitive
Theory. Please let me know whether this is possible.

I may be contacted via e-mail at wfritz@nnu.edu, by mail at ███████████
████████████████████████████████ or by telephone at ████████████████

Thank you for your contributions to our work in schools. We have made and
will make much progress in helping students and teachers by using your
constructs as a foundation for motivating good learning and teaching.

Best regards,

Bill Fritz
Doctoral Student at Northwest Nazarene University, Nampa, ID
Assistant Superintendent, School District of Indian River County, Vero
Beach, FL

Bandura, A. (1986).* Social foundations of thought and action: A social
cognitive theory. *Englewood Cliffs, NJ: Prentice-Hall.

From: bandura@stanford.edu Albert Bandura
To:
Date: Mon, 13 Jan 2014 17:34:03 -0500
Subject: RE: Permission Request to Use Triadic Reciprocality Chart

Permission granted to use the diagram.

Albert Bandura

Appendix C

Marzano Learning Map and Permission

Marzano Art and Science of Teaching Teacher Evaluation Model

DOMAIN 1: CLASSROOM STRATEGIES AND BEHAVIORS

Lesson Segments Involving Routine Events

Learning Goals & Feedback
What will I do to establish and communicate learning goals, track student progress, and celebrate success?
• Providing Clear Learning Goals and Scales to Measure those Goals
• Tracking Student Progress
• Celebrating Student Success

Rules & Procedures
What will I do to establish or maintain classroom rules and procedures?
• Establishing Classroom Routines
• Organizing Physical Layout of the Classroom for Learning

Lesson Segments Addressing Content

Interacting With New Knowledge
What will I do to help students effectively interact with the new knowledge?
• Identifying Critical Information
• Organizing Students to Interact with New Knowledge
• Previewing New Content
• Chunking Content into "Digestible Bites"
• Processing of New Information
• Elaborating on New Information
• Recording and Representing Knowledge
• Reflecting on Learning

Practicing & Deepening Knowledge
What will I do to help students practice and deepen their understanding of new knowledge?
• Reviewing Content
• Organizing Students to Practice and Deepen Knowledge
• Using Homework
• Examining Similarities and Differences
• Examining Errors in Reasoning
• Practicing Skills, Strategies, and Processes
• Revising Knowledge

Generating & Testing Hypotheses
What will I do to help students generate and test hypotheses about new knowledge?
• Organizing Students for Cognitively Complex Tasks
• Engaging Students in Cognitively Complex Tasks Involving Hypothesis Generating and Testing
• Providing Resources and Guidance

Lesson Segments Enacted on the Spot

Student Engagement
What will I do to engage students?
• Noticing and Reacting when Students are Not Engaged
• Using Academic Games
• Managing Response Rates
• Using Physical Movement
• Maintaining a Lively Pace
• Demonstrating Intensity and Enthusiasm
• Using Friendly Controversy
• Providing Opportunities for Students to Talk about Themselves
• Presenting Unusual or Intriguing Information

Adherence to Rules & Procedures
What will I do to recognize and acknowledge adherence and lack of adherence to classroom rules and procedures?
• Demonstrating "Withitness"
• Applying Consequences
• Acknowledging Adherence to Rules and Procedures

Teacher/Student Relationships
What will I do to establish and maintain effective relationships with students?
• Understanding Students' Interests and Backgrounds
• Using Behaviors that Indicate Affection for Students
• Displaying Objectivity and Control

High Expectations
What will I do to communicate high expectations for all students?
• Demonstrating Value and Respect for Low Expectancy Students
• Asking Questions of Low Expectancy Students
• Probing Incorrect Answers with Low Expectancy Students

Domain 1 identifies the 41 key strategies revealed by research for effective teaching presented in a robust, easy-to-understand model of instruction based on the Art and Science of Teaching.

All 41 Key Strategies are organized into 9 Design Questions, which are further organized into 3 Lesson Segments.

Learning Sciences International®
LEARNING AND PERFORMANCE MANAGEMENT

Marzano Art and Science of Teaching Teacher Evaluation Model

DOMAIN 2: PLANNING AND PREPARING

Planning and Preparing for Lessons and Units
1. Effective Scaffolding of Information within Lessons
2. Lessons within Units
3. Attention to Established Content Standards

Planning and Preparing for Use of Resources and Technology
1. Use of Available Traditional Resources
2. Use of Available Technology

Planning and Preparing for Special Needs of Students
1. Needs of English Language Learners
2. Needs of Special Education Students
3. Needs of Students Who Lack Support for Schooling

DOMAIN 3: REFLECTING ON TEACHING

Evaluating Personal Performance
1. Identifying Areas of Pedagogical Strength and Weakness
2. Evaluating the Effectiveness of Individual Lessons and Units
3. Evaluating the Effectiveness of Specific Pedagogical Strategies and Behaviors

Developing and Implementing a Professional Growth Plan
1. Developing a Written Growth and Development Plan
2. Monitoring Progress Relative to the Professional Growth and Development Plan

DOMAIN 4: COLLEGIALITY AND PROFESSIONALISM

Promoting a Positive Environment
1. Promoting Positive Interactions with Colleagues
2. Promoting Positive Interactions about Students and Parents

Promoting Exchange of Ideas and Strategies
1. Seeking Mentorship for Areas of Need or Interest
2. Mentoring Other Teachers and Sharing Ideas and Strategies

Promoting District and School Development
1. Adhering to District and School Rules and Procedures
2. Participating in District and School Initiatives

Learning Sciences International®
LEARNING AND PERFORMANCE MANAGEMENT

Hi Bill,

You are able to use the learning map as long as the copyright is included on it. Please let me know if you have any other questions.

Thank you,

Lindsey N. Devers, Ph.D.
Senior Research Analyst
LearningSciencesInternational
1400 Centrepark Boulevard | Suite 1000
West Palm Beach, FL 33401
Email: ldevers@learningsciences.com

Hi Bill,

Kathy Marx forwarded me your email. Let's schedule a time to talk towards the end of the week. What times are you available on Friday?

Best Regards,

Lindsey N. Devers, Ph.D.
Senior Research Analyst

LearningSciencesInternational
1400 Centrepark Boulevard | Suite 1000
West Palm Beach, FL 33401
Email: ldevers@learningsciences.com

Dear Ms. Marx:

Thank you for your perseverance in making contact with me.

As I mentioned on the phone, I am conducting a Doctoral study identifying whether relationships exist between teachers' self-efficacy beliefs and Marzano observation ratings. My research is under the supervision of Northwest Nazarene University in Nampa, ID.

One element that I need assistance with is whether Marzano ratings (Innovating, Applying, Developing, Beginning, Not Using) are best identified as ordinal variables or nominal variables. An opinion from LSI psychometricians would be appreciated. Additionally, if you can provide a sampling of citations for studies where Marzano ratings are treated as a dependent variable, that would help too...APA format is preferred.

Lastly, please reply to verify permission to include the "Learning Map" as an appendix to my dissertation so that readers may better understand the evaluation domains and elements. Your suggested citation would be helpful too.

Thank you.

Bill Fritz
Doctoral Candidate
Northwest Nazarene University
Nampa, Idaho

Appendix D

Site Authorization to Conduct Study

School District of

March 25, 2014

William T. Fritz,

Dear Mr. Fritz:

I hereby grant permission for you, William T. Fritz, to conduct the study entitled, "A Quantitative Study Investigating the Efficacy Beliefs of Elementary Mathematics Teachers," within the School District during the 2014-15 School year. I also grant permission for to act as a surrogate to administer the survey, gather the data, and provide it in a manner whereby the identity of subjects is unknown.

I understand that you will be surveying fourth and fifth grade teachers within our school system, and acknowledge that we will provide 2012-13 VAM data and evaluation data to you, as described in your request letter.

I verify that I am authorized to grant this permission on behalf of the School District of , Florida.

Sincerely,

Superintendent

Equal Opportunity Educator and Employer

Information redacted to maintain confidentiality of subjects.

Appendix E

Informed Consent Form

TEACHER BELIEFS SURVEY
INFORMATION AND INFORMED CONSENT

A. PURPOSE AND BACKGROUND

The researcher for this study is a Doctoral student. The study investigates teachers' beliefs about their capabilities to help students learn. A specific focus is mathematics teaching beliefs. The study will also investigate possible relationships between teacher beliefs and teacher performance evaluation ratings. We appreciate your involvement in helping us investigate how to better serve the needs of students and teachers.

Due to the sensitive nature of the data, an individual within the school district, Bruce Green, will be gathering the data and providing it to the researcher so that all personally identifiable information is removed.

You are being asked to participate in this study because you are a healthy volunteer, over the age of 18.

B. PROCEDURES

If you agree to be in the study, the following will occur:

You are being asked to electronically sign this Informed Consent Form, volunteering to participate in the study.

You will complete an electronic survey related to your teaching beliefs. This survey will also include a few demographic questions.

Your recent evaluation results and Florida value-added (VAM) scores will be extracted as part of a data file, and coded (to maintain your confidentiality).

Your survey results will be coded so that they may be "linked" to your performance evaluation in a confidential matter, with your name unknown to Mr. Green or the researcher.

These procedures will be competed electronically within a time period during the fall of 2014 decided upon by the investigator. The survey will take a total time period of about 20 minutes.

C. RISKS/DISCOMFORTS

It is possible that some of the discussion questions may make you uncomfortable or upset. While this is unlikely, you are free to decline to answer any questions you do not wish to answer or to stop participation at any time.

For this research project, the researchers are requesting demographic information. Due to the make-up of Florida's population, it is possible that the combined answers to these questions will make an individual person identifiable. The researchers will make every effort to protect your confidentiality. However, if you are uncomfortable answering any of these questions, you may leave them blank.

Confidentiality: Participation in research may involve a loss of privacy; however, your records will be handled as confidentially as possible. No individual identities will be used in any reports or publications that may result from this study. All data from will be kept in a locked file cabinet at the premises of the researcher and the key to the cabinet will be kept in a separate location. In compliance with the Federalwide Assurance Code, data from this study will be kept for three years, after which all data from the study will be destroyed (45 CFR 46.117).

D. BENEFITS

There will be no direct benefit to you from participating in this study. However, the information you provide may help educators to better understand the belief structures of teachers and how these beliefs are derived, as well as possible relationships with classroom practice.

E. PAYMENTS

There are no payments for participating in this study.

However, as an incentive for participation, at the conclusion of the study, a drawing will be held (using code numbers) and two individuals from the study will be identified to receive a $50.00 gift card for Publix Supermarket.

The School District will ensure that the winners receive the gift cards since the researcher will not know the identity of participants.

F. QUESTIONS

If you have questions or concerns about participation in this study, you should first talk with the surrogate investigator, Bruce Green. He can be contacted via email at bruce.green@indianriverschools.org, via telephone at 772-564-3099 or by writing: 1990 25th Street, Vero Beach, FL 32960.

Should you feel distressed due to participation in this, you should contact your own health care provider.

G. CONSENT

You may make a copy of this consent form to keep.

PARTICIPATION IN RESEARCH IS VOLUNTARY. You are free to decline to be in this study, or to withdraw from it at any point. Your decision as to whether or not to participate in this study will have no influence on your present or future status as a student at any University.

By clicking continue, I give my consent to participate in this study.

 THE UNIVERSITY HUMAN RESEARCH REVIEW COMMITTEE HAS REVIEWED THIS PROJECT FOR THE PROTECTION OF HUMAN PARTICIPANTS IN RESEARCH.

Appendix F

Reminder Message to Participants

Dear Study Participant:

Last month, you agreed to participate in a study related to teacher beliefs and their relationship with student and teacher performance. This message is a friendly reminder to click the following link to complete the survey. The total time it will take you to assist is no more than 20 minutes, and by completing the survey, you will help advance our educational practice.

If you have questions about the study or survey, you are welcome to contact me at
_____ or by phone at _____

Thank you again for considering participation in this study.

Best regards,

_____
Research Liaison
School District of ▮▮▮▮▮▮▮▮▮▮▮▮

(link to survey)

Appendix G

Survey Instruments

Teachers' Sense of Efficacy Scale[1] (short form)

Teacher Beliefs	**How much can you do?**								
Directions: This questionnaire is designed to help us gain a better understanding of the kinds of things that create difficulties for teachers in their school activities. Please indicate your opinion about each of the statements below. Your answers are confidential.	Nothing		Very Little		Some Influence		Quite A Bit		A Great Deal
1. How much can you do to control disruptive behavior in the classroom?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2. How much can you do to motivate students who show low interest in school work?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
3. How much can you do to get students to believe they can do well in school work?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
4. How much can you do to help your students value learning?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
5. To what extent can you craft good questions for your students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
6. How much can you do to get children to follow classroom rules?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
7. How much can you do to calm a student who is disruptive or noisy?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
8. How well can you establish a classroom management system with each group of students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
9. How much can you use a variety of assessment strategies?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
10. To what extent can you provide an alternative explanation or example when students are confused?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
11. How much can you assist families in helping their children do well in school?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
12. How well can you implement alternative strategies in your classroom?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

Directions for Scoring the Teachers' Sense of Efficacy Scale[1]

Developers: Megan Tschannen-Moran, College of William and Mary
Anita Woolfolk Hoy, the Ohio State University.

Construct Validity

For information the construct validity of the Teachers' Sense of Teacher efficacy Scale, see:

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education, 17,* 783-805.

Factor Analysis

It is important to conduct a factor analysis to determine how your participants respond to the questions. We have consistently found three moderately correlated factors: *Efficacy in Student Engagement, Efficacy in Instructional Practices*, and *Efficacy in Classroom Management*, but at times the make up of the scales varies slightly. With preservice teachers we recommend that the full 24-item scale (or 12-item short form) be used, because the factor structure often is less distinct for these respondents.

Subscale Scores

To determine the *Efficacy in Student Engagement, Efficacy in Instructional Practices*, and *Efficacy in Classroom Management* subscale scores, we compute unweighted means of the items that load on each factor. Generally these groupings are:

Long Form
Efficacy in Student Engagement:	Items 1, 2, 4, 6, 9, 12, 14, 22
Efficacy in Instructional Strategies:	Items 7, 10, 11, 17, 18, 20, 23, 24
Efficacy in Classroom Management:	Items 3, 5, 8, 13, 15, 16, 19, 21

Short Form
Efficacy in Student Engagement:	Items 2, 3, 4, 11
Efficacy in Instructional Strategies:	Items 5, 9, 10, 12
Efficacy in Classroom Management:	Items 1, 6, 7, 8

Mathematics Teaching Efficacy Beliefs Instrument (MTEBI)

Developed by Enochs, Smith, and Huinker, (2000), used with permission

Please indicate the degree to which you agree or disagree with each statement that follows by **clicking the appropriate description to the right.**

	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
1. When a student does better than usual in mathematics, it is often because the teacher exerted a little extra effort. Mathematics	A	B	C	D	E
2. I will continually find better ways to teach mathematics.	A	B	C	D	E
3. Even if I try very hard, I do not teach mathematics as well as I do most subjects.	A	B	C	D	E
4. When the mathematics grades of students improve it is often due to their teacher having found a more effective teaching approach.	A	B	C	D	E
5. I know the steps necessary to teach mathematics concepts effectively.	A	B	C	D	E
6. I am not very effective in monitoring mathematics activities.	A	B	C	D	E
7. If students are underachieving in mathematics, it is most likely due to ineffective mathematics teaching.	A	B	C	D	E
8. I generally teach mathematics ineffectively.	A	B	C	D	E

	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
9. The inadequacy of *a student's mathematics* background can be overcome by good teaching.	A	B	C	D	E
10. The low mathematics achievement of some students cannot generally be blamed on their teacher.	A	B	C	D	E
11. When a low-achieving child progresses in mathematics, it is usually due to extra attention given by the teacher.	A	B	C	D	E
12. I understa*nd mathematics concepts well* enough to be effective in teaching mathematics.	A	B	C	D	E
13. Increased effort in mathematics teaching produces little change in some students' mathematics achievement.	A	B	C	D	E
14. The teacher is generally responsible for the achievement of students in mathematics.	A	B	C	D	E
15. Students' achievement in mathematics is directly related to their teacher's effectiveness in mathematics teaching.	A	B	C	D	E
16. If parents comment that their child is showing more interest in mathematics at school, it is probably due to the performance of the child's teacher.	A	B	C	D	E
17. I find it difficult to use manipulatives to explain to students why mathematics works.	A	B	C	D	E
18. I am typically able to answer students' mathematical Questions.	A	B	C	D	E
19. I wonder if I have the necessary skills to teach mathematics.	A	B	C	D	E

	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
20. Given a choice, I would not invite my principal to evaluate my mathematics teaching.	A	B	C	D	E
21. When a student has difficulty understanding a mathematics concept, I am usually at a loss as to how to help the student understand it better.	A	B	C	D	E
22. When teaching mathematics, I usually welcome student questions.	A	B	C	D	E
23. I do not know what to do to turn students on to mathematics.	A	B	C	D	E

Demographic Survey

Please provide some information about yourself.

1. Age in years _____
2. Number of years teaching _____
3. Gender M   F
4. Please identify *all* mathematics classes you successfully completed while in college (identify any that apply):

   Mathematics teaching methods course
   College Algebra
   College Trigonometry
   Pre-Calculus or Math Analysis
   Statistics
   Calculus
   Remedial College Mathematics Coursework (at a pre-college level) to meet prerequisites
   Geometry
   Discrete Mathematics
   Other (please describe) _____

5. Please indicate the number of mathematics instruction professional development hours in which you have participated during the last five years.

   _____   21+ hours
   _____   16-20 hours
   _____   11-15 hours
   _____   6-10 hours
   _____   1-5 hours
   _____   no mathematics specific professional development

6. Ethnicity:
   African-American/Black _____
   Latino/Hispanic _____
   Native American _____
   Asian / Pacific Islander _____
   Caucasian/White _____
   Multi-Racial _____
   Other _____

Appendix H

Raw Likert Data

Teacher Sense of Efficacy Scale
Short Form
Response Tallies

Rating Denotations:
1: Nothing   3: Very Little   5: Some Influence   7: Quite a Bit   9: A Great Deal

Q1 How much can you do to control disruptive behavior in the classroom?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	1	1	7	8	15
Percent	0.0%	0.0%	0.0%	0.0%	3.1%	3.1%	21.9%	25.0%	46.9%

Q2 How much can you do to motivate students who show low interest in school work?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	2	3	14	5	8
Percent	0.0%	0.0%	0.0%	0.0%	6.3%	9.4%	43.8%	15.6%	25.0%

Q3 How much can you do to calm a student who is disruptive or noisy?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	1	2	12	7	10
Percent	0.0%	0.0%	0.0%	0.0%	3.1%	6.3%	37.5%	21.9%	31.3%

Q4 How much can you do to help your students value learning?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	1	0	4	4	6	17
Percent	0.0%	0.0%	0.0%	3.1%	0.0%	12.5%	12.5%	18.8%	53.1%

Q5 To what extent can you craft good questions for your students?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	1	0	10	8	13
Percent	0.0%	0.0%	0.0%	0.0%	3.1%	0.0%	31.3%	25.0%	10.6%

Q6 How much can you do to get children to follow classroom rules?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	0	2	4	8	18
Percent	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	12.5%	25.0%	56.3%

Q7 How much can you do to get students to believe they can do well in school work?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	1	2	6	8	15
Percent	0.0%	0.0%	0.0%	0.0%	3.1%	6.3%	18.8%	25.0%	46.9%

Q8 How well can you establish a classroom management system with each group of students?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	0	0	4	10	18
Percent	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	12.5%	31.3%	56.5%

Q9 How much can you use a variety of assessment strategies?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	2	6	10	6	8
Percent	0.0%	0.0%	0.0%	0.0%	6.3%	18.8%	31.3%	18.8%	25.0%

Q10 To what extent can you provide an alternative explanation or example when students are confused?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	0	2	7	10	13
Percent	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	21.9%	31.3%	40.6%

Q11 How much can you assist families in helping their children do well in school?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	5	7	7	7	6
Percent	0.0%	0.0%	0.0%	0.0%	15.6%	21.9%	21.9%	21.9%	18.8%

Q12 How well can you implement alternative strategies in your classroom?

Rating	1	2	3	4	5	6	7	8	9
Frequency	0	0	0	0	2	4	8	9	9
Percent	0.0%	0.0%	0.0%	0.0%	6.3%	12.5%	25.0%	28.1%	28.1%

Mathematics Teaching Efficacy Beliefs Instrument
Response Tallies

SD = Strongly Disagree
D = Disagree
A = Agree
SA = Strongly Agree

Note: (R) denotes reverse coded question.

1. When a student does better than usual in mathematics, it is often because the teacher exerted a little extra effort.

Rating	SD	D	Uncertain	A	SA
Frequency	0	69	5	15	6
Percent	0.0%	18.8%	15.6%	46.9%	18.8%

2. I will continually find better ways to teach mathematics.

Rating	SD	D	Uncertain	A	SA
Frequency	0	0	0	8	24
Percent	0.0%	0.0%	0.0%	24.2%	72.7%

3. Even if I try very hard, I do not teach mathematics as well as I do most subjects. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	17	10	2	3	0
Percent	53.1%	31.3%	6.3%	9.4%	0.0%

4. When the mathematics grades of students   improve it is often due to their teacher having found a more effective teaching approach.

Rating	SD	D	Uncertain	A	SA
Frequency	0	0	3	20	9
Percent	0.0%	0.0%	9.4%	62.50%	28.1%

5. I know the steps necessary to teach mathematics concepts effectively.

Rating	SD	D	Uncertain	A	SA
Frequency	0	0	4	15	13
Percent	0.0%	0.0%	12.5%	46.9%	40.6%

6. I am not very effective in monitoring mathematics activities. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	13	14	2	2	0
Percent	41.9%	45.2%	6.5%	6.5%	0.0%

7. If students are underachieving in mathematics, it is most likely due to ineffective mathematics teaching.

Rating	SD	D	Uncertain	A	SA
Frequency	1	9	6	16	0
Percent	3.1%	28.1%	18.8%	50.0%	0.0%

8. I generally teach mathematics ineffectively. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	19	9	2	1	0
Percent	61.3%	29.0%	6.5%	3.2%	0.0%

9. The inadequacy of *a student's mathematics* background can be overcome by good teaching.

Rating	SD	D	Uncertain	A	SA
Frequency	0	2	7	21	2
Percent	0.0%	6.3%	21.9%	65.6%	6.3%

10. The low mathematics achievement of some students cannot generally be blamed on their teacher. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	2	13	8	9	0
Percent	6.3%	40.6%	25.0%	28.1%	0.0%

11. When a low-achieving child progresses in mathematics, it is usually due to extra attention given by the teacher.

Rating	SD	D	Uncertain	A	SA
Frequency	0	2	7	18	5
Percent	0.0%	6.3%	21.9%	56.3%	15.6%

12. I understa*nd mathematics concepts well* enough to be effective in teaching mathematics.

Rating	SD	D	Uncertain	A	SA
Frequency	0	0	1	11	19
Percent	0.0%	0.0%	3.2%	35.5%	61.3%

13. Increased effort in mathematics teaching produces little change in some students' mathematics achievement. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	4	20	4	3	1
Percent	12.5%	62.5%	12.5%	9.4%	3.1%

14. The teacher is generally responsible for the achievement of students in mathematics.

Rating	SD	D	Uncertain	A	SA
Frequency	0	3	2	23	4
Percent	0.0%	9.4%	6.3%	71.9%	12.5%

15. Students' achievement in mathematics is directly related to their teacher's effectiveness in mathematics teaching.

Rating	SD	D	Uncertain	A	SA
Frequency	0	5	4	18	5
Percent	0.0%	15.6%	12.5%	56.3%	15.6%

16. If parents comment that their child is showing more interest in mathematics at school, it is probably due to the performance of the child's teacher.

Rating	SD	D	Uncertain	A	SA
Frequency	0	2	8	19	3
Percent	0.0%	6.3%	25.0%	59.4%	9.4%

17. I find it difficult to use manipulatives to explain to students why mathematics works. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	9	17	2	4	0
Percent	28.1%	53.1%	6.3%	12.5%	0.0%

18. I am typically able to answer students' mathematical questions.

Rating	SD	D	Uncertain	A	SA
Frequency	0	0	1	12	19
Percent	0.0%	0.0%	3.1%	37.5%	59.4%

19. I wonder if I have the necessary skills to teach mathematics. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	13	16	0	2	1
Percent	39.4%	48.5%	0.0%	6.3%	3.1%

20. Given a choice, I would not invite the principal to evaluate my mathematics teaching. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	16	10	1	3	1
Percent	51.6%	32.3%	3.2%	9.7%	3.2%

21. When a student has difficulty understanding a mathematics concept, I am usually at a loss as to how to help the student understand it better. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	17	12	2	1	0
Percent	51.5%	36.4%	6.3%	3.1%	0.0%

22. When teaching mathematics, I usually welcome student questions.

Rating	SD	D	Uncertain	A	SA
Frequency	0	0	0	15	17
Percent	0.0%	0.0%	0.0%	46.9%	53.1%

23. I do not know what to do to turn students on to mathematics. (R)

Rating	SD	D	Uncertain	A	SA
Frequency	11	17	2	2	0
Percent	34.4%	53.1%	6.3%	6.3%	0.0%

Appendix I

Permission for Use of Teachers' Sense of Efficacy Scale

COLLEGE OF
EDUCATION AND
HUMAN ECOLOGY
THE OHIO STATE UNIVERSITY

ANITA WOOLFOLK HOY, PH.D.

PROFESSOR
PSYCHOLOGICAL STUDIES IN EDUCATION

Dear William Fritz:

You have my permission to use the *Teachers' Sense of Efficacy Scale* in your research. A copy of both the long and short forms of the instrument as well as scoring instructions can be found at:

http://www.coe.ohio-state.edu/ahoy/researchinstruments.htm

Best wishes in your work,

Anita Woolfolk Hoy

Anita Woolfolk Hoy, Ph.D.
Professor

Appendix J

Permission for Use of Mathematics Teacher Efficacy Beliefs Instrument

```
From: wfritz@nnu.edu William Fritz
To:
Date: Sun, 29 Dec 2013 07:25:41 -0500
Subject: Permission to use MTEBI
```

Dear Dr. Huinker:

I am a doctoral student at Northwest Nazarene University (Nampa, ID) and am interested in using the Mathematics Teaching Efficacy Beliefs Instrument to gather data pertaining to my dissertation. The topic of my dissertation is mathematics teaching efficacy beliefs of 3rd grade teachers, and relationships between these beliefs and classroom performance using both the Marzano teacher evaluation system as well as State of Florida Value Added Model scores.

I am in the initial planning stages now, and would like to request your permission to utilize the instrument. Please let me know whether this is possible or whether there is a process I must follow to procure such permission.

The instrument is germane to my study due to its content specificity.

I may be contacted via e-mail at wfritz@nnu.edu, by mail at ▉▉▉▉▉▉ ▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉▉ or by telephone at ▉▉▉▉▉▉▉▉

Please also let me know whether you would be willing to informally advise on my study via phone conversation or electronic communication. With the advent of high-stakes evaluation systems and growth model systems in our state (Florida) this study seems especially relevant.

Also, would you know how to reach Dr. Phillip L. Smith. He is no longer listed on the University of Wisconsin - Milwaukee website and it appears that that there are a couple of other people out there with the same name. Any assistance you can provide would be helpful.

Thank you.

Best regards,

Bill Fritz

From: huinker@uwm.edu DeAnn Huinker

To: wfritz@nnu.edu William Fritz
Date: Fri, 3 Jan 2014 11:06:36 -0500
Subject: Re: Permission to use MTEBI

Bill,

Certainly you have our permission to use the MTEBI in your research. Please consider this email as verification of our permission.

Dr. Smith and Dr. Enochs have retired and I do not have current contact information for either of them.

If you have questions regarding the instrument, I can try to address them. Email is usually the best way to reach me, however if you are in need of a phone conversation, we could arrange that as well.

Article reference for the MTEBI:
Enochs, L., Smith, P. And Huinker, D. "Establishing factorial validity of the Mathematics Teaching Efficacy Beliefs Instrument." School Science and Mathematics, 2000,100 (4),194-202.
Best to you in your research,
DeAnn Huinker

From: wfritz@nnu.edu William Fritz
To:
Date: Sat, 28 Dec 2013 07:13:54 -0500
Subject: Permission for use of MTEBI

Dear Dr. Enochs:

I am a doctoral student at Northwest Nazarene University (Nampa, ID) and am interested in using the Mathematics Teaching Efficacy Beliefs Instrument to gather data pertaining to my dissertation. The topic of my dissertation is mathematics teaching efficacy beliefs of 3rd grade teachers, and relationships between these beliefs and classroom performance using both the Marzano teacher evaluation system as well as State of Florida Value Added Model scores.

I am in the initial planning stages now, and would like to request your permission to utilize the instrument. Please let me know whether this is possible or whether there is a process I must follow to procure such permission.

The instrument is germane to my study due to its content specificity.

I may be contacted via e-mail at wfritz@nnu.edu, by mail at ███████ ████████████████████████████████ or by telephone at ████████████

Please also let me know whether you would be willing to informally advise on my study via phone conversation or electronic communication. With the advent of high-stakes evaluation systems and growth model systems in our state (Florida) this study seems especially relevant.

Thank you.

Best regards,

Bill Fritz

From: enochsl@onid.orst.edu LARRYENOCHS

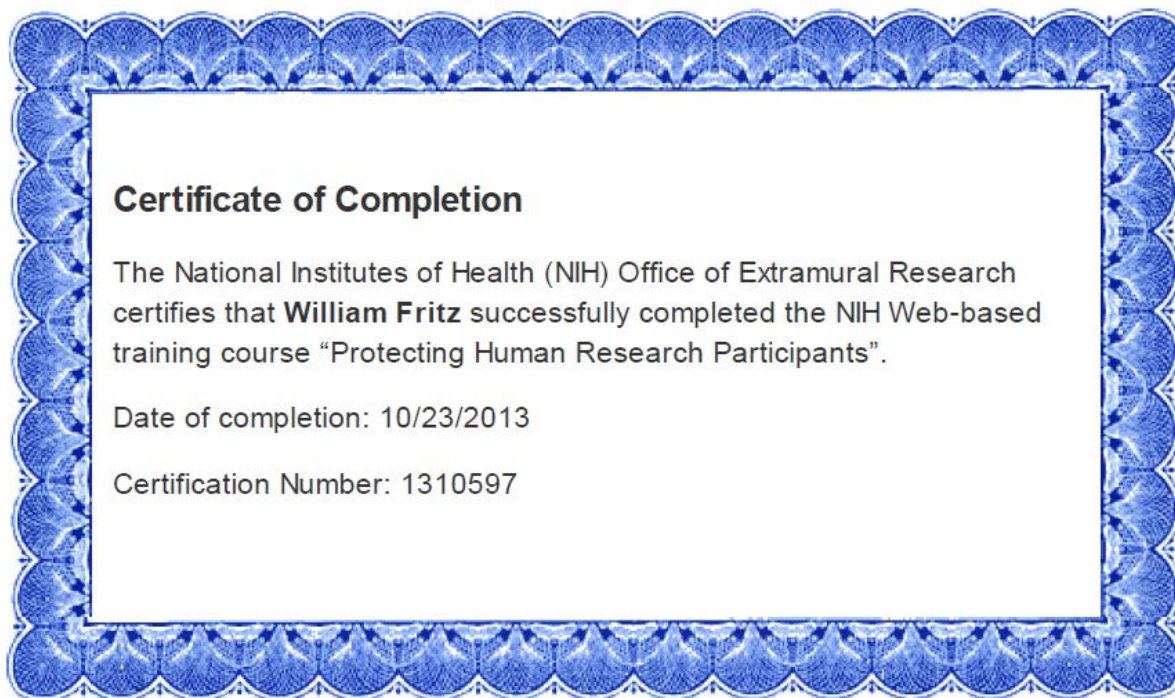To:
Date: Sun, 29 Dec 2013 02:50:59 -0500
Subject: Re: Permission for use of MTEBI

You certainly may use it. Good luck

Appendix K

Protecting Human Research Certificate

Protecting Human Subject Research Participants

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **William Fritz** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 10/23/2013

Certification Number: 1310597

Appendix L

Permission to Utilize Florida Value-Added Technical Report
(Florida Department of Education, 2011)

From: Gaitanis, Jason [mailto:Jason.Gaitanis@fldoe.org]
Sent: Monday, August 18, 2014 12:00 PM
To: Fritz, William
Subject: RE: Request to utilize and cite technical report

Bill,

Yes, you can go ahead and use it. I have confirmed with AIR that this is the final version.

Thanks,

Jason

From: Fritz, William
Sent: Sunday, August 17, 2014 5:32 PM
To: Gaitanis, Jason
Subject: FW: Request to utilize and cite technical report

Dear Mr. Gaitanis:

I have not heard back from you since our last correspondence on August 4, 2014. Although the footer on the table of contents page indicates that the report is a draft and not to be cited, is it permissible to use and cite the report, as explained in the correspondence below?

Bill Fritz

From: Fritz, William
Sent: Monday, August 04, 2014 5:13 PM
To: 'Gaitanis, Jason'
Subject: RE: Request to utilize and cite technical report

The footer on the table of contents page

From: Gaitanis, Jason [mailto:Jason.Gaitanis@fldoe.org]
Sent: Monday, August 04, 2014 5:10 PM
To: Fritz, William
Subject: RE: Request to utilize and cite technical report

Where are you seeing working draft on the document?

From: Fritz, William
Sent: Monday, August 04, 2014 5:08 PM
To: Gaitanis, Jason
Subject: RE: Request to utilize and cite technical report

Dear Mr. Gaitanis

The Technical report for 12-13, accessed from this link, still says it is a "working draft." May I cite that manual?

Bill Fritz

From: Gaitanis, Jason [mailto:Jason.Gaitanis@fldoe.org]
Sent: Monday, August 04, 2014 4:58 PM
To: Fritz, William
Cc: Copa, Juan
Subject: RE: Request to utilize and cite technical report

Mr. Fritz,

The final versions of the FCAT VAM Model Technical reports for 11-12 and 12-13 can be found here on the Department's website:

http://www.fldoe.org/profdev/studentgrowth.asp

Sincerely,

Jason Gaitanis
Policy, Research and Accountability Coordinator
Division of Accountability, Research and Measurement
Florida Department of Education
Turlington Building, Suite 544
325 West Gaines Street
Tallahassee, FL 32399
(850) 245 – 0411
jason.gaitanis@fldoe.org

Begin forwarded message:

From: "Fritz, William"
Date: August 2, 2014 at 4:19:13 PM EDT
To: "Juan.Copa@fldoe.org" <Juan.Copa@fldoe.org>
Subject: Request to utilize and cite technical report

Dear Deputy Commissioner Copa:

I am currently engaged in Doctoral work at Northwest Nazarene University. My dissertation topic includes Florida value-added scores as one of my variables. I would like to use information from the technical report published by the Florida Department of Education, in conjunction with American Institutes for Research, as part of my dissertation, but see that it is a working draft, which is not to be cited or distributed. The report link is:

https://www.fldoe.org/committees/doc/Value-Added-Model-Technical-Report.docx

Can you please verify whether there was a draft of this document finalized, or whether there is some more current document that explains the value added mechanics? If not, how would I go about gaining permission to use and cite this document?

Thank you for your assistance with this matter.

Best regards,

Bill Fritz