STARS

Electronic Theses and Dissertations, 2004-2019

2014

An Analysis of the Variables in Implementation of the Marzano Causal Teacher Evaluation System in the State of Florida

Matthew Phillips
University of Central Florida

STARS Citation

Showcase of Text, Archives, Research & Scholarship

AN ANALYSIS OF THE VARIABLES IN IMPLEMENTATION OF THE MARZANO
CAUSAL TEACHER EVALUATION SYSTEM IN THE STATE OF FLORIDA

by

MATTHEW A. PHILLIPS
B.S. University of Central Florida, 1997
M.Ed. University of Central Florida, 2005

A dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Education
in the School of Teaching, Learning, and Leadership in the College of Education and
Human Performance at the University of Central Florida
Orlando, Florida

Summer Term
2014

Major Professor: Barbara A. Murray

ABSTRACT

The focus of this research was on the different ways in which the Marzano causal teacher evaluation system was implemented in the state of Florida during the 2013-2014 school year. Teacher evaluation systems were compared based on seven implementation variables. The researcher then sought to determine what relationships may or may not exist between districts that followed the research-based implementation variables and districts that did not.

It was determined that implementation of the Marzano causal evaluation system varied widely. In six of the seven implementation variables, it was determined that there were statistically significant relationships in teacher performance ratings between districts that followed the research-based implementation variables and districts that did not.

This means that the Marzano causal teacher evaluation system should be used as it was intended--following research based best practice. The research in this study showed that when implementation variables were altered in the Marzano causal teacher evaluation system, teacher performance ratings were less accurate.

In 2009, my mom passed away and although she will not be standing next to me on graduation day, I know she will be watching--proud of my accomplishment. From the time I was little, she was always my biggest cheerleader. My mom and dad told me--I could do it! When I finished my master's degree, I told mom that I was one day going to go back and earn my doctoral degree. Her pride was my motivation, my drive, to completing my doctorate.

I dedicate this dissertation to my parents, I love you mom and dad!

ACKNOWLEDGMENTS

Completing the doctoral program at the University of Central Florida has been an amazing journey. I have grown personally and professionally and feel I am a better person as a result of my journey. My growth and success was only possible with the support of my family, friends, and university faculty.

To my family, I could not have done this without your support. Thank you to my wife and son for giving up family time on many stressful days. Evan and Monica, love you both with all my heart.

To my friends, thank you for all of the kind words of encouragement you have provided me on the most difficult days. Beverly, thank you for being there with an ear to bend and the right words of encouragement. Something you said to me that helped me on tough days was "You can do this, you are a finisher!" Thank you for being a friend!

The faculty in my University of Central Florida program were second to none. Thank you to my dissertation committee for the constant input and guidance through this process. In particular, Dr. Barbara Murray has been an amazing teacher and mentor for me throughout my masters and doctoral programs at the University of Central Florida.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1
THE PROBLEM AND ITS CLARIFYING COMPONENTS

Background of the Study

In 2009, the Race to the Top federal grant was authorized. A total of 40 states, including the District of Columbia, applied for the initial round of funding for this federal grant (U.S. Department of Education, 2009). The stated purpose of this grant was to promote educational change and innovation. The Race to the Top grant was comprised of six sections: Great Teachers and Leaders, State Success Factors, Standards and Assessments, General Selection Criteria, Turning Around the Lowest-Achieving Schools, and Data Systems to Support Instruction. In order to receive money from this federal grant, state and school district educational agencies wrote grant applications addressing improvements in the six areas. Under the first section, Great Teachers and Leaders, a significant emphasis was put on "improving teacher and principal effectiveness based on performance" (U.S. Department of Education, 2009, p. 7). The specifics of this improvement goal called for new systems to be instituted to evaluate teachers. By 2012, 36 states and the District of Columbia had authorized legislation to implement new teacher evaluation systems (National Council on Teacher Quality, 2012).

According to a national policy review of teacher evaluations by the National Council on Teacher Quality (2012) there have been two major changes that have taken place in teacher evaluations since 2009. The first was the implementation of a student data component in teacher evaluations, where teacher evaluations were to be determined

in part based on student academic performance on standardized assessments. The second major change was the implementation of new teacher observation systems.

Since 2009, the state of Florida has sought to receive Race to the Top funding. As a result, Florida has implemented significant changes to teacher evaluation systems. Florida Senate Bill 736, called the Student Success Act, was signed into law in 2011. This bill required changes in the way teachers were evaluated. It outlined specific characteristics teacher evaluation systems must include, e.g., rubrics for teacher performance levels, and the evaluation of teachers based on their implementation of research-based instructional strategies. Additionally, the bill required that all teacher evaluation systems be reviewed and approved by the Florida Department of Education.

Prior to this bill there was one state adopted model for teacher observations, the Florida Performance Measurement System (FPMS). Additionally, local districts were allowed to bargain other teacher observation protocols such as narrative observations with local unions. Prior to the Student Success Act there was no Florida Department of Education review or approval required for district-used teacher observation and evaluation systems.

By the 2012-2013 school year, every county in the state of Florida had implemented new teacher evaluation systems. There are 67 counties in the state of Florida. Of these, 61% used the Marzano causal teacher evaluation system, 24% used the Charlotte Danielson framework for teaching, and 16% developed their own systems for teacher evaluations. These data show that in the state of Florida the Marzano causal teacher evaluation system was the most widely adopted teacher evaluation system.

The purpose of the Marzano causal teacher evaluation system has been to accurately evaluate teachers. The accuracy of the data provided from the Marzano causal teacher evaluation system has been critical for school leaders so that they may make appropriate staffing decisions and provide accurate feedback to teachers on their instructional practice. Though there is a research-based way to implement the Marzano causal teacher evaluation system (Livingston & Livingston, 2012; Marzano, Frontier, & Livingston, 2011), the implementation of new evaluation systems has been left to local districts to determine and bargain with local unions. This means that implementation of the Marzano causal teacher evaluation system may vary among different school districts. Thus, it is possible that changing implementation variables could alter the accuracy of teacher performance ratings and diminish the overall stated purpose of the evaluation system. This study sought to examine what relationships, if any, exist between teacher performance ratings and different implementation variables.

<u>Statement of the Problem</u>

A total of 25 Florida school districts have adopted the Marzano causal teacher evaluation system. Implementation of the Marzano causal teacher evaluation system varies among the respective 25 school districts. To date, no research has analyzed the different ways in which school districts have implemented the Marzano causal teacher evaluation system. Additionally, no research has been conducted to see how variations in implementation might affect teacher performance ratings.

Purpose of the Study

The purpose of the study was to review Florida school district implementation plans of the Marzano causal teacher evaluation system and to report the different ways in which school districts have implemented the system. The Marzano causal teacher evaluation system was designed to accurately evaluate teachers and provide feedback to teachers. Changing implementation variables of the system might affect the reliability or validity of teacher instructional feedback and performance ratings.

By understanding how these variables possibly affect teacher performance ratings and feedback, educational leaders should be able to better implement the Marzano causal teacher evaluation system. This study was conducted with the intention of assisting educational leaders in knowing what the implementation variables were and how these variables may or may not relate to different teacher performance ratings. If certain variables were found to impact teacher performance ratings, educational leaders could avoid altering these variables when implementing the system.

Theoretical Framework

Fredrick Winslow Taylor (1911), in his theory of scientific management, introduced the idea that there were scientifically best ways in which to do work most effectively. Taylor outlined four principals of scientific management: (a) developing a science for work, (b) scientifically selecting and training employees, (c) ensuring that the work being done followed scientific best practice, and (d) planning and overseeing of workers by managers. Furthermore, in a 1911 congressional hearing Taylor stated that,

4

"A proper day's work is a matter of accurate, careful scientific investigation" (Taylor, 1911, p. 14). Influenced by the work of Fredrick Taylor, evaluation systems were developed to measure and monitor worker productivity. These measurements were used to evaluate employee effectiveness, to determine employee wages, and to justify the termination of struggling employees. The goal was to improve job productivity by ensuring the employment of the most efficient employees. In determining employee wages, it was believed that an incentive pay system based on measurable job performance would motivate employees to work harder. By using scientifically measurable job performance standards, struggling employees could be identified correctly, and terminated. Under this system, it was believed that the threat of termination would motivate employees to work harder. If struggling employees were terminated, better employees could be hired to replace inefficient workers.

Scientific management provided the theoretical framework on which performance-related employee evaluation systems would be based. These evaluation systems have been called performance appraisals or performance evaluations. Performance appraisals identify scientifically-proven methods and skills that an employee should be utilizing. They follow a systematic process to evaluate employees based on these scientifically-proven methods and skills.

The Marzano causal teacher evaluation system is a performance appraisal system which draws upon Taylor's scientific management theory, evaluating teachers based on their use of standardized, scientifically proven instructional strategies. This study was

conducted to determine if the Marzano causal teacher evaluation system was being implemented in a way that ensures the accuracy of the evaluation system.

Definition of Terms

Classroom walkthrough--A brief classroom visit and evaluation usually lasting between three and eight minutes. These visits are usually not scheduled, and are conducted unannounced.

Formal observation--An observation that is prearranged and scheduled with a teacher. The administrator conducts a pre- and post-conference with the teacher and observes the teacher for an entire lesson or class period using the Marzano causal teacher evaluation system protocol for Domain 1.

Informal observation--An unannounced, unscheduled observation during which the administrator evaluates all or part of a lesson or class period using the Marzano causal teacher evaluation system protocol for Domain 1.

Marzano causal teacher evaluation system--A new evaluation model in the state of Florida that includes teacher observation protocols. Teachers are evaluated based on four Domains. Domain 1 is based on classroom observations of teacher use of instructional strategies and behaviors. Domain 2 is based on planning and preparing. Domain 3 is an evaluation of a teacher's ability to reflecting on teaching. Domain 4 is an evaluation of a teacher's collegiality and professionalism.

Teacher performance rating--A summative evaluation score assigned to a teacher in the state of Florida based on a value added model (VAM). Using student VAM data

and a teacher instructional practice score, a summative teacher performance rating

(highly effective, effective, needs improvement, and unsatisfactory) is assigned. For

teachers working less than three years, the needs improvement rating is replaced with the

classification of developing.

Value added model (VAM)--A method for determining a teacher's effectiveness

based on student growth data.

<div align="center">Delimitations of Study</div>

The following Florida school districts were included in this study: Bradford,

Broward, Calhoun, Charlotte, Collier, Franklin, Gadsden, Gilchrist, Indian River,

Jackson, Lafayette, Lake, Leon, Martin, Nassau, Orange, Osceola, Pasco, Putnam, Santa

Rosa, Seminole, St. Johns, St. Lucie, Union, and Palm Beach County. This list represents

the Florida school districts that implemented the Marzano causal teacher evaluation

system during the 2012-2013 school year.

This study was focused on the implementation of the Marzano causal teacher

evaluation system in 25 school districts across the state of Florida and used data reported

by the Florida Department of Education (FLDOE). These data included: school district

approved instructional performance evaluation system plans and district reported teacher

performance data for the 2012-2013 school year. The dependent variable was delimited

to reported teacher performance ratings.

The independent variables were delimited to the following seven implementation

variables:

1. Implementation of teacher deliberate practice.

2. Number of formal observations conducted.

3. The scope of instructional elements observed during classroom observations.

4. Implementation of pre- and post-conferencing during the formal observation cycle.

5. Implementation of informal observations.

6. Implementation of classroom walkthroughs.

7. Implementation of Domain 2, Domain 3 and Domain 4 teacher performance ratings.

Limitations of Study

One limitation of the study was that only implementation variables reported to the state under Race to the Top requirements were identified. There may have been other variables, not reported, which could have affected teacher instructional practice scores. For example, districts did not indicate how much training stakeholders received on the new system. This variable could have greatly impacted how effectively a district implemented the Marzano system, thereby altering teacher instructional practice scores.

Another limitation of this study was that teachers' student growth scores were part of summative teacher performance ratings. Similar to the implementation of teacher observation systems used to determine instructional practice scores, local districts had the ability to alter how student growth data impacted teacher performance ratings. The variation in implementation of how student growth scores impacted teacher performance

ratings could have caused a statistical relationship among different district teacher performance ratings.

<div align="center">Research Questions</div>

1. What are the different ways in which 25 Florida school districts implemented the Marzano causal teacher evaluation system during the 2012-2013 school year?

2. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to deliberate practice?

   H_{02}--There is no relationship in the teacher performance ratings of school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to deliberate practice.

3. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regards to the number of formal observations conducted?

   H_{03}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the number of formal observations conducted.

4. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations?

H_{04}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations.

5. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing?

   H_{05}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing.

6. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regards to the implementation of informal observations?

   H_{06}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of informal observations.

7. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs?

   H_{07}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs.

8. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to implementation of Domain 2, Domain 3, and Domain 4 evaluations?

H_{08}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of Domain 2, Domain 3,, and Domain 4 evaluations.

Methodology

Population and Sample

The population for this study consisted of the 67 public school districts in Florida. The following Florida school districts comprised the sample of schools that were included in this study: Bradford, Broward, Calhoun, Charlotte, Collier, Franklin, Gadsden, Gilchrist, Indian River, Jackson, Lafayette, Lake, Leon, Martin, Nassau, Orange, Osceola, Pasco, Putnam, Santa Rosa, Seminole, St. Johns, St. Lucie, Union, and Palm Beach County. This list represents the Florida school districts that implemented the Marzano causal teacher evaluation system during the 2012-2013 school year.

Data Collection and Analysis

First, approval was obtained from the University of Central Florida Internal Review Board to conduct this research study (Appendix A). The next step was to

research the Marzano causal teacher evaluation system. Through this research, as reported in the literature review in Chapter 2, a model of how to implement the Marzano causal teacher evaluation system was established.

All districts in the state of Florida reported their 2012-2013 teacher evaluation plans to the Florida Department of Education (FLDOE). This information was accessible on the FLDOE (2014a) Educator Recruitment, Development, & Retention webpage. This webpage reported all district evaluation systems and included the tools and procedures every district used to evaluate instructional staff. The researcher reviewed this webpage and identified school districts using the Marzano causal teacher evaluation system. Then, those district implementation plans were read, analyzed and compared to the literature on best practice. Based on the comparison of best practice and actual district implementation variables were established. These variables represent ways in which the Florida school districts changed the Marzano causal teacher evaluation system from the best practice model. The implementation variables, as identified by the researcher, were: (a) implementation of teacher deliberate practice; (b) number of formal observationsconducted; (c) scope of instructional elements observed during classroom observations; (d) implementation of pre- and post-conferencing during the formal observation cycle; (e) implementation of informal observations; (f) implementation of classroom walkthroughs; and (g) implementation of Domain 2, Domain 3 and Domain 4 in teacher performance ratings.

As each of the district implementation plans were read and analyzed narrative notes were taken. These notes were used in writing narrative descriptions of how each

district implemented the Marzano causal teacher evaluation system during the 2012-2013 school year. District names were removed, and districts were assigned a random letter in the narrative.

Districts implementing the Marzano causal teacher evaluation system were listed, by the random letter on a data collection sheet with implementation variables identified in separate columns. If a district changed an implementation variable from the research based model, it was noted on the data collection sheet as "N" indicating the implementation model had not been followed. If the variable was implemented according to the research based model, the data collection sheet was marked with a "Y".

District-wide teacher performance ratings were obtained from the FLDOE (2014b) Educator Recruitment, Development, & Retention webpage. The percentages of teachers identified as highly effective, effective, needs improvement/developing and unsatisfactory were identified for each school district. This information was added to the data collection sheet. The categories for needs improvement/developing and unsatisfactory were combined because there were so few teachers classified as unsatisfactory that this category was statistically insignificant.

The data collection sheet was then imported into IBM SPSS analytical predictive software (SPSS) to run separate Chi-square tests for Research Questions 2-8 to determine if there were statistically significant relationships between the implementation variables and teacher performance ratings.

Organization of the Study

This dissertation consists of five chapters. Chapter 1 has presented an introduction, statement of problem, purpose of study, background of the study, theoretical framework, definitions of terms, delimitations of study, limitations of study, research questions, and the methodology of the study. Chapter 2 contains a review of relevant literature related to the study. Chapter 3 presents the study methodology which includes an introduction, population, selection of the sample, research design and rationale, collection of data, an analysis of the data, and a summary. Chapter 4 contains the results of the data analysis. This includes an introduction, descriptive statistics, data analysis, summary and a conclusion. Chapter 5, the final chapter in the dissertation, consists of an introduction, summary of the study, discussion of the findings, implications for practice, recommendations for further research, conclusions and a summary.

CHAPTER 2
REVIEW OF LITERATURE

Introduction

The first use of the term, performance appraisals, is unknown, but the modern use

of the term is thought to have come into use shortly after the end of World War II (Wiese

& Buckley, 1998). Performance appraisal systems evaluate employee performance based

on established standards, criteria, data and goals. It is human nature to judge and

evaluate, performance appraisal systems add structure to this process. Performance

appraisal systems serve several purposes. First, they help to identify effective an

ineffective workers. With performance appraisal systems, an employee's deficiencies

can be identified and corrected through training (Maasa & Reddy, 2009). Performance

appraisal systems are designed to provide ongoing feedback to employees to improve job

performance, and feedback is found to be a highly effective strategy in improving

behavior and teaching new behaviors (Hattie, 2009). It is important to understand

performance appraisal systems in relation to the present study, because the Marzano

causal teacher evaluation system is a performance appraisal system for teachers.

The Marzano causal teacher evaluation system combines research from both

supervision and evaluation; thus, it has two purposes. As an evaluation system, the

purpose is to accurately evaluate teachers so that effective and ineffective teachers can be

identified. This information can be used to make retention, dismissal, and staffing

decisions. The second purpose of this performance appraisal system is to improve teacher

pedagogy by establishing a clinical supervision model to provide accurate feedback to teachers.

<center>Teacher Performance Appraisal Systems</center>

Cubberly (1929) was responsible for a very early application of scientific management to schools. He outlined how scientific management could be applied to schools, school operations, and teacher supervision, noting specific examples of what a supervisor should look for when visiting classrooms and what feedback to give teachers in regards to specific scientifically based instructional strategies when observing instruction.

From 1927 to 1932, the Hawthorn Studies, under the direction of Mayo (1946), expanded the role of what would become employee performance appraisal systems. Mayo found, in these studies, that factors other than pay and the threat of being fired motivated employees. One factor found to improve employee performance was feedback between a worker and supervisor (Mayo, 1946). This developed into the idea that performance appraisal systems could be more than ways to determine wages and justify the termination of struggling employees. They could provide employees with feedback based on measurable standards. This feedback, in turn, would help employees improve their job performance.

Another educational innovation that impacted teacher performance appraisals was clinical supervision. A Harvard professor, Cogan, developed one of the first systems of clinical supervision while working with the Master of Arts program (Reavis, 1976). This

new model was similar to supervisory practices used in the medical field. In this model, supervisor and practitioner were involved in a series of observations, conferences, and discussions. Goldhammer (1969) developed a significant and widely adopted model of this clinical supervision process. In the Goldhammer model, the following five step process to the supervision of teachers was created: (a.) pre-observation conference, (b) classroom observation, (c) analysis, (d) a supervision conference, and (e) analysis of the analysis. This model had a profound impact on teacher evaluation systems, as many modern teacher observation systems have continued to incorporate parts of this process. The Marzano causal teacher evaluation system also used a variation of the clinical supervision model.

Specific teacher instructional strategies and student behaviors to look for during observations were not established in the Goldhammer model. During an observation, using Goldhammer's model, the observer was to collect descriptive and comprehensive data on what they saw and heard (Goldhammer, 1969). During the pre-conference, the observer and teacher could agree on specific data to look for, but no specific strategies to look for or implement were established in Goldhammer's work.

Later, Hunter (1980) expanded on the clinical supervision model to include specific instructional behaviors and characteristics that a supervisor should look for during observations. In her model, strategies such as clear objectives, instructional hooks, and guided practice were identified as effective instructional strategies teachers should use, and that observers should look for during instruction.

17

In the 1980s, a new development in the teacher supervision process, the differentiation of teacher evaluations based on individual needs, emerged. As Glatthorn (1984) described in his text on differentiated supervision, teachers should have input about their improvement needs. He proposed the idea that clinical supervision was not the only model to improve and coach teachers, suggesting that teachers should have input as to a variety of supervisory models that could be used to supervise them. These models were: clinical supervision, cooperative supervision, self-directed supervision and administrative monitoring.

The Marzano causal teacher evaluation system was developed from the research base on teacher supervision and evaluation. This system is a performance appraisal system used to both evaluate and supervise teachers. As an evaluation system, it needs to accurately determine teacher effectiveness. As a supervisory tool, it needs to provide a structure for the clinical supervision process so as to provide teachers accurate feedback on their instructional practice.

Because a performance evaluation system serves as both an evaluation tool and supervision model there can be a conflict of interest (Sergiovanni & Starratt, 2007). A conflict arises because with the goal of supervision, a trusting relationship is required. When teachers know that the supervisor coaching and providing feedback is also evaluating them, the relationships can be strained. Teachers, realizing that as an evaluator the supervisor may have to dismiss them, may be less likely to be open and share instructional challenges with their supervisor.

A performance evaluation used as a supervisory tool should provide effective feedback to teachers to improve their instructional pedagogy. In 1992, Hattie found that feedback was one of the most powerful influencing factors on achievement. Kluger and DeNisi (1996) found in their meta-study that feedback was most effective when low levels of threat were perceived by the learner, because low threat levels allow attention to be paid to the feedback. In the Marzano causal teacher evaluation system, the clinical supervision process is also part of the evaluation process. This means that feedback might be perceived as high stakes as it is directly related to teacher evaluation ratings, staffing decisions, and possible merit pay in the state of Florida. High stakes teacher feedback on instruction is counter to the goal of teacher improvement. This is a stark example of how the supervision process and evaluation process can be at odds with one another.

Grubb (2007), in his reappraisal of performance appraisal systems, determined that these systems failed to improve employee skills. This may be explained in other studies where employee perceptions and understanding of performance appraisal systems were considered to be a large factor in determining how effective performance appraisal systems would be in the work place (Cardy & Dobbins, 1994). The authors of one popular study concluded that, if done well, performance appraisal systems can effectively evaluate and be helpful to the supervision processes (Lawler, Benson, & McDemmot, 2012). These studies showed that performance appraisal system effectiveness had varied, and that effectiveness was dependent on how the system had been implemented.

New Teacher Evaluation Systems

Teacher evaluations provide summative judgments of teacher ability (Kosmoski, 2000). With 21st century modifications in teacher evaluations, led by Race to the Top, there has been an assumption by politicians and the public that previous teacher evaluations failed to accurately measure teacher instructional practice. This perception of failing teacher evaluations was supported in the results of Farkas, Johnson, and Duffett's 2003 survey. These researchers found that 77% of teachers believed that teacher evaluations were ineffective and that administrators simply went through the motions to complete the annual evaluation process. What is not clear in the survey results is whether these perceptions were related to poor implementation and administrator practice or the evaluation systems themselves.

Four reoccurring trends in changes to new teacher evaluation systems have been identified in state policy, national policy, and educational literature. Those four changes are: (a) the inclusion of student performance data in teacher evaluations, (b) an increase in frequency of teacher observations, (c) better identification of struggling teachers and (d) better performance feedback data to teachers on their instructional pedagogy.

The first change in teacher evaluation systems is the inclusion of student performance data in teacher evaluations. The score on a teacher's evaluation calculated from student assessment data is called the student growth score. The student growth score is a score ranging from 1 to 4 based on student academic growth as measured by the FCAT or other standardized student assessment. Although districts were free to develop their own methods for calculating teacher growth scores, the state of Florida

approved model for this calculation was the Value Added Model (VAM). The score, determined through teacher observations, is called the instructional practice rating. Although districts are free to develop their own methods for determining teacher instructional practice scores, the state-approved model for this, is the Marzano causal teacher evaluation system.

In the state of Florida, the student growth score and instructional practice scores have been determined independent of one another. After a teachers' student growth score and instructional practice scores are determined, they are averaged together to determine a summative end-of-year teacher performance rating. Final teacher performance ratings are then divided into one of four performance categories: Highly Effective, Effective, Needs Improvement or Developing (for teachers working less than three years), and Unsatisfactory.

The second change for teacher evaluation systems has been to increase the frequency of teacher observations. Prior to the implementation of the Race to the Top grant, according to the results of a study by the National Council on Teaching Quality (2010), teacher evaluations took place infrequently across the entire United States. It was found that teacher observations, as part of the evaluation process, took place either twice a year, once a year, or once every two years. In Florida, the previous generation observation protocol, the Florida Performance Measurement System (FPMS), was administered once or twice a year. This infrequent teacher evaluation feedback was a problem with previous generation teacher observation protocols (Farkas et al., 2003; Kersten & Israel, 2005; Rothman & Toch, 2008).

The third change for teacher evaluation systems has been to better identify struggling teachers. In a large 2004 study, Haefele analyzed teacher evaluation systems in multiple states. It was found that 99% of teachers were highly effective on end-of-year evaluations. The conclusion from the study was that the evaluation systems failed to identify poor performing teachers, and that there was a trend of teacher evaluation inflation. Not addressed in the study was the reason(s) for teacher evaluation inflation. The researchers speculated that this condition could have resulted from any one of a number of reasons, e.g., improper implementation of the system and not a result of the systems themselves. New teacher evaluation systems have been implemented with the intention of reducing teacher evaluation inflation and to have a system to more accurately identify struggling teachers.

Prior to the Florida Senate Bill 736, the Student Success Act, there was one state-adopted model for teacher observations called the Florida Performance Measurement System (FPMS). An example of the FPMS observation form is included in Appendix B. Additionally, local districts were allowed to implement other observation tools, like narrative observations, that could be used to evaluate teachers.

The FPMS was a data collection tool designed to be used in conjunction with the observation of teachers and provide summative feedback to teachers on their instructional practice. Unlike the Marzano causal teacher evaluation system, the FPMS was not designed to assign a rating to a teacher (Micceri, Peterson, & Smith, 1985). The FPMS used a checklist format where the frequency at which effective and ineffective instructional strategies was recorded. A requirement of new teacher evaluation systems

22

in the state of Florida, as outlined by Senate Bill 736, was that they assign a summative rating to teachers based on a range of proficiency scales. These rating scales outlined levels of performance for different observed teaching strategies. By adding proficiency scales for the use of instructional strategies, one is able to differentiate between highly effective, effective, and struggling teacher performance levels. Rather than concentrating on the frequency of use for instructional strategies, the evaluator focuses on how well a particular instructional strategy is being implemented. In a checklist format, these differentiations are much more difficult to discern (Rothman & Toch, 2008). The newer observation protocols have been directly linked to a teacher's end-of-year performance rating. The scores from observations conducted throughout the year are used in calculating a teacher's end-of-year evaluation performance rating.

When using the Marzano causal teacher evaluation system, the observer rates the use of instructional strategies on a proficiency scale ranging from 1-4 where a rating of 1 would be considered beginning use of a strategy and a rating of 4 would be considered highly effective (innovative) use of a strategy. The scores are totaled, and the teacher is given a final evaluation rating based on the observations.

In 2008, Rothman and Toch conducted a national review of state evaluation systems. They found that other states had similar teacher evaluation systems that did not rate teachers on a proficiency scale of implementation during observations. These other systems were similar to the FPMS and did not assign teacher ratings based on their implementation of instructional strategies.

With goals to include the use of student performance data in teacher evaluations, to increase the frequency of teacher observations, to better identify struggling teachers and to provide better performance feedback data to teachers, new teacher evaluation systems have been implemented to replace previous generation systems. Although the new teacher evaluation systems are early in implementation, some research has been initiated to analyze potential problems.

In 2005, Kersten and Israel conducted a research study on next generation evaluation systems implemented as a result of Race to the Top. They surveyed principals who were implementing new evaluation systems. The principals surveyed believed that the new observation systems were complicated and cumbersome and that because of this, the systems were misunderstood and misused. As a result of this, teachers and teacher unions were resistant to implementing the new systems even though they agreed the systems had improved over past systems. Because the new systems were cumbersome, the principals surveyed indicated that they spent more time evaluating teachers and were not able to devote an appropriate amount of time to other areas of instructional leadership.

The Marzano causal teacher evaluation system, having been used only a few years in the state of Florida at the time of the present study, was early in its implementation. Thus, little research has been conducted on its effectiveness. Specifically, no research has been conducted to investigate whether the system does a better job of identifying struggling teachers or if it has done a better job providing feedback to teachers to improve their instructional practice. A 2013 study by Flowers (2013) yielded no

correlation between student achievement and teacher performance ratings in the Marzano system.

A possible concern with the Marzano system could be in the varied levels of implementation among districts. Research has been conducted to indicate best practice and proper ways to implement the system; however, every local school district in Florida has to bargain with local teacher unions as to how teacher evaluation systems will be implemented. This leaves room for misuse in implementation. Changing and altering originally intended implementation strategies could cause teacher evaluations to be inaccurate.

The Marzano Causal Teacher Evaluation System Implementation Model

The Marzano causal teacher evaluation system is a performance appraisal system. It requires student performance outcomes to be combined with observable and measurable teacher behaviors to determine a teacher instructional practice score. The Marzano causal teacher evaluation system was first introduced, explained, and defined by Marzano, Frontier, and Livingston in their 2000 book. The system they described was a supervision model. This supervision model was then adopted by the state of Florida Department of Education as the state-approved teacher evaluation system. No research or studies were found in this literature review that tested the validity and reliability of the Marzano system.

The Marzano causal teacher evaluation system is divided into four domains. Domain 1 is Classroom Strategies and Behaviors. This Domain focuses on classroom

instruction. Of the four domains, Domain 1 has the largest and most direct impact on student learning. Within this domain there are 41 instructional elements that impact student learning. Teachers are observed during both formal and informal observations on the 41 elements. A teacher's use of the elements is rated according to the following proficiency scale: Innovating (4), Applying (3), Developing (2), Beginning (1), and Not Using (0). For each element, the scale is different. The rating of Innovative is the highest use of the element where Beginning indicates the use of the element is missing or in error. A score of Not Using would indicate the element was called for but was not used by the teacher. It is important to understand that even though there are 41 elements, not all of them are applicable during a teacher observation. Only the elements being used or called for should be evaluated during a classroom observation. During a typical 45-minute lesson, a range of some elements will be observed.

Erickson, Krampe and Tesch-Romer (1993), defined deliberate practice as a process for professionals to refine and improve their practice. Applied to the Marzano causal teacher evaluation system, this would be the process of teachers identifying specific instructional elements upon which to focus improvement. Based on prior observation feedback, these elements should be ones that the teacher wants and needs to improve. In this process, the teacher creates a professional development plan to improve in this area, and the supervisor agrees to observe and provide feedback on the teacher's use of this instructional element. The purpose of this process is teacher improvement, and the evaluation score is based on the teacher's ability to improve in this area.

The Teacher Development Tool Kit (Livingston & Livingston, 2012) created by the Marzano Research Laboratories outlines the minimum frequency for which formal observations should take place. Teachers should be divided into two groups of teachers, initial status teachers (beginning teachers) and professional status (experience teachers). Initial status teachers are recommended to have two formal observations by the middle of the school year and two more by the end of the school year. Professional status teachers should have one formal observation by the middle of the year and one formal observation by the end of the year.

The process for formal observations includes an administrator conducting both pre- and post-conferences with teachers. The pre-conference is conducted prior to the observation to gather information about the observation, identify key elements to be observed, and to review the teacher's lesson. During the post conference, the supervisor ratings are discussed and student evidence may be reviewed. Teacher evaluation ratings can be adjusted after the post-conference.

Informal observations are unannounced and unscheduled, and an administrator observes part or all of a classroom lesson, providing evaluative feedback to a teacher. Teachers should have a frequent number of these throughout the year (Livingston & Livingston, 2012), and no maximum should be place on the number of informal observations to be conducted in classrooms.

Another best practice outlined in the Teacher Development Tool Kit (Livingston & Livingston, 2012) is the inclusion of classroom walkthroughs in teacher evaluations.

Classroom walkthroughs are brief class visits where a couple elements are observed, and the observer provides evaluative feedback to the teacher during this brief visit.

Domain 2 is Planning and Preparation. This Domain has eight elements that are related to teacher planning and preparation. Domain 3 is Reflecting on Teaching. Teachers are evaluated based on their ability to reflect on their teaching. Domain 4 is collegiality and professionalism. Teachers are evaluated in this domain based on professional practice. Within these domains, teachers are evaluated on different elements within each domain using the same levels from Domain 1 (innovative, applying, developing, beginning and not using). Although the level names are the same, the proficiency scales for each element in each domain are different. There is no research to suggest how often teachers should be evaluated in their use of Domains 2, 3, and 4. Researchers have simply noted that teachers should be evaluated and receive feedback on their usage of the elements within these domains

<u>Implementation Variables</u>

Based on the review of literature, seven implementation variables were identified. These represent best practice in the implementation of the Marzano causal teacher evaluation system

Deliberate Practice

With the Marzano causal teacher evaluation system, teachers identify targeted areas for improvement. Teachers select an instructional element in Domain 1 of the Marzano causal teacher evaluation system on which to focus their professional

development. This element is observed by an evaluator and feedback is given to teachers

on their improved use of this element. The growth, or improvement, of this element is

included in teachers' instructional practice scores. This deliberate practice gives teachers

input into their professional development. Often this deliberate practice element is used

in conjunction with an Individual Professional Development Plan created by the teacher.

Number of Formal Observations

This refers to the number of formal observations conducted on teachers during the

year. For an initial status teacher, two formal observations by mid-year and two

observations by the end of the year are recommended. For a professional status teacher,

one formal observation by mid-year and one formal observation by the end of year should

be conducted (Livingston & Livingston, 2012).

Scope of Formal Elements Observed

During a formal observation, the observer evaluates the elements being

implemented or elements that are called for but not being implemented. These are the

elements evaluated during the observation (Livingston & Livingston, 2012). This would

indicate that of the 41 classroom instructional elements, any could be observed during an

observation and that no limit should be placed on which elements should be observed.

Typically, for a formal observation most of the elements are planned for and discussed

during the pre-conference. However, there are elements that Marzano (2011) has

described as being enacted on the spot that could be observed but not planned for.

Pre/Post Conferencing

The formal observation process is part of the clinical supervision model as established by Goldhammer (1969). This process calls for supervisors and teachers to meet before and after an observation. The purpose of pre-conferencing is to identify the observation tool, focus the observation on specific teaching elements, help the teacher plan and refine the lesson, and help the observer start collecting data on the observation (Marzano, 2011).

After the observation, a post-conference is conducted between the observer and teacher. The purpose of this post-conference is to provide feedback to the teacher and for the observer to ask clarifying questions before finalizing the evaluation score (Marzano, 2011).

Informal Observations

Informal observations should be included in teacher evaluations. During an informal observation, the evaluator appears in a classroom unannounced and unscheduled to observe a lesson. During informal observations, the observer evaluates the elements being implemented or elements that are called for, but not being implemented (Livingston & Livingston, 2012). This would indicate any of the 41 classroom instructional elements could be evaluated during an informal observation and that no limit should be placed on which elements to evaluate. There is no maximum or minimum number of informal observations to be conducted. According to Livingston and Livingston (2012), the amount of informal observations should be numerous.

Classroom Walkthroughs

Classroom walkthroughs should be included in teacher evaluations. During a classroom walkthrough, the evaluator comes to a classroom unannounced and unscheduled to observe class instruction briefly. The observer evaluates the elements being implemented or elements that are called for, but not being implemented (Livingston & Livingston, 2012). This would indicate any of the 41 classroom instructional elements could be evaluated during a classroom walkthrough and that no limit should be placed on which elements could be evaluated. There is no research to suggest an appropriate amount of classroom walkthroughs to include in teacher performance ratings.

Domains 2, 3 and 4 Evaluations

Domains 2, 3 and 4 should be included in teacher evaluations. Domain 2 refers to the teacher's ability to plan and prepare for instruction. In this domain, there are eight specific elements that can be evaluated. Domain 2 should be part of a teacher's overall instructional practice score. However, this domain might count less toward the instructional practice score than Domain 1 of classroom strategies and behaviors.

Domain 3 refers to the teacher's ability to reflect in teaching. In this domain, there are five specific elements that can be evaluated. Domain 3 should be part of a teacher's overall instructional practice score. However, this domain might count less toward the instructional practice score than Domain 1.

Domain 4 refers to a teacher's collegiality and professionalism. In this Domain, there are six specific elements that can be evaluated. Domain 4 should be part of a teacher's overall instructional practice score. However, this domain might count less toward the instructional practice score than Domain 1.

CHAPTER 3
METHODOLOGY

Introduction

The purpose of this study was to research the different ways in which school

districts in the state of Florida have implemented the Marzano causal teacher evaluation

system and to determine if there were statistically significant relationships between

implementation variables and teacher performance ratings. This chapter contains a

detailed explanation of the methods used to conduct the research and answer the research

questions. It has been organized into the following sections: (a) research questions (b)

population (c) selection of the sample (d) research design and rationale (e) collection of

data (f) analysis of data and (g) summary.

Research Questions

1. What are the different ways in which 25 Florida school districts implemented

the Marzano causal teacher evaluation system during the 2012-2013 school

year?

2. What relationship, if any, exists in teacher performance ratings between

school districts that followed and did not follow the Marzano causal teacher

evaluation system in regard to deliberate practice?

H_{02}--There is no relationship in the teacher performance ratings of school

districts that followed and did not follow the Marzano causal teacher

evaluation system in regard to deliberate practice.

3. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regards to the number of formal observations conducted? H_{03}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the number of formal observations conducted.

4. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations?

   H_{04}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations.

5. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing?

   H_{05}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing.

6. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regards to the implementation of informal observations? H_{06}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of informal observations.

7. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs? H_{07}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs.

8. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to implementation of Domain 2, Domain 3, and Domain 4 evaluations? H_{08}--There is no relationship in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of Domain 2, Domain 3,, and Domain 4 evaluations.

Population and Sample

The population for the study was comprised of the 67 public school districts in the state of Florida. This study was conducted on a sample of 25 school districts that implemented the Marzano causal teacher evaluation system during the 2012-2013 school year in the state of Florida. The school districts included in this study were: Bradford, Broward, Calhoun, Collier, Franklin, Gadsden, Gilchrist, Indian River, Jackson, Lafayette, Lake, Leon, Martin, Nassau, Orange, Osceola, Pasco, Putnam, Santa Rosa, Seminole, St. Johns, St. Lucie, Union, and Palm Beach Counties.

Selection of the Sample

No random sampling methods were used to select the school districts used in the study. All school districts in the State of Florida that met the research study criteria were used in the study. All of the school districts identified were those in the state of Florida that used the Marzano causal teacher evaluation system during the 2012-2013 school year, reported their implementation procedures to the FLDOE, and reported teacher performance ratings to the FLDOE. The school districts using the Marzano causal teacher evaluation system were identified from reported information on the FLDOE (2014a) website. School districts reporting their Marzano causal teacher evaluation systems were identified from published data on the FLDOE (2014a) website on the Educator Recruitment, Development, & Retention webpage. School districts reporting teacher performance data were also identified from published data on the FLDOE (2014b) website.

Research Design and Rationale

Based on the research questions, this non-experimental study used a mixed-methods approach to research. Based on the questions, several different research designs were required. Table 1 shows the research questions, type of tests used, the independent variables, and the dependent variables for each research question.

Table 1

Research Questions, Statistical Tests, and Variables

Research Questions	Test	Independent Variable	Dependent Variable
1. What are the different ways in which 25 Florida school districts implemented the Marzano causal teacher evaluation system during the 2012-2013 school year?	Qualitative Research		
	For Research Questions 2-8		
What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow Marzano causal teacher evaluation system in regard to: 2. deliberate practice? 3. the number of formal observations conducted? 4. the scope of elements observed during formal observations? 5. the implementation of pre- and post-conferencing? 6. the implementation of informal observations? 7. the implementation of classroom walkthroughs? 8. the implementation of Domain 2, Domain 3, and Domain 4 evaluations?	Chi-square test	Following or not following the model	Teacher performance ratings

For Research Question 1, the researcher was required to collect district implementation documentation. This information was analyzed and compared to the implementation model described in Chapter 2. To answer this question a qualitative approach was used to describe the implementation of each district in detailed narratives.

Research Questions 2-8 sought to determine what relationships, if any, existed between the seven implementation variables and teacher performance ratings. Each question was analyzed separately using a Chi-square test. The Chi-square test was used because it is the most appropriate statistical test when both the independent and dependent variables are categorical. The independent variable was categorical data with two categories: districts that followed (y) and districts that did not follow (n) the implementation model of best practice. The dependent variables were categorical data, with three teacher evaluation categories where numbers of teachers in each category were reported.

Collection of Data

Collection of the data were initiated using the FLDOE website and making a list of school districts that used the Marzano causal teacher evaluation system for teacher evaluations during the 2012-2013 school year. These school districts were determined using the FLDOE (2014a) webpage identifying the teacher evaluation systems each school district was using. Using these data, identified districts were listed on a data collection sheet (See Appendix C).

The researcher relied on the research-based implementation model that had been established for the Marzano causal teacher evaluation system and was described in the literature review for the present research. This model describes the intended way in which the Marzano causal teacher evaluation system should be implemented, i.e., seven implementation variables. The implementation variables were also listed on the data collection sheet.

District implementation plans for the Marzano causal teacher evaluation system which had been retrieved from the FLDOE (2014d) were read, analyzed, and compared to the research-based implementation model. Notes were taken and a narrative description for each district was written. The 25 school districts implementing the Marzano causal teacher evaluation system were each assigned a random letter. This letter was used to represent the school district on the data collection sheet and in the written narrative description of the district implementation. If a district followed the implementation model, a "y" was recorded on the data collection sheet for that district. If a district altered one of the implementation variables, it was recorded on the data collection sheet as an "n" for not following the model.

Teacher performance ratings for each district were obtained from the FLDOE 2014b). The ratings were reported as the total number of teachers in the school district identified as: highly effective, effective, needs improvement or developing (three or less years of experience), and unsatisfactory. Because the category of unsatisfactory had fewer than 5 teachers identified in each district, the category was statistically insignificant. Therefore, the researcher combined two categories: needs improvement or

developing (three or less years of experience) and unsatisfactory indicating a teacher is not meeting performance expectations. The numbers of teachers in each of the three categories were converted into percentages, and the percentages were recorded for each district on the data collection sheet.

The data collection sheet was subsequently imported into SPSS. The data imported included district letters, implementation variables, and district teacher performance ratings. The implementation variables were coded with a "y" if the district followed the research-based protocol and "n" if the district did not follow the research based protocol.

Analysis of Data

Research Question 1

Research Question 1 sought to identify the different ways in which 25 Florida school districts implemented the Marzano causal teacher evaluation system during the 2012-2013 school year. As reported in Chapter 2, a review of literature was conducted and a research-based implementation model was established for the Marzano causal teacher evaluation system. This model is considered the intended way in which the Marzano causal teacher evaluation system was to be implemented. Based on this model, seven implementation variables were established. The implementation variables were: implementation of teacher deliberate practice, (b) number of formal observations conducted, (c) the scope of instructional elements observed during classroom observations, (d) implementation of pre- and post-conferencing during the formal

observation cycle, (e) implementation of informal observations, (f) implementation of classroom walkthroughs, and (g) implementation of Domain 2, Domain 3, and Domain 4 in teacher performance ratings

The district implementation plans were read and analyzed by the researcher. The different ways in which the 25 school districts implemented the Marzano causal teacher evaluation system were compared to the research-based implementation model and the nine identified implementation variables. A decision was made by the researcher based on the written documentation whether or not a district implemented the particular variable in accordance to the research based model. In situations where a district was found not to be implementing the system according to the research based model, an explanation and justification was stated in the narrative description. In the narrative description, district names were removed and districts were assigned a random letter designation. After the narrative description was written the district letters, the seven implementation variables and district teacher evaluation ratings were added to the data collection sheet (Appendix C).

Research Question 2

Research Question 2 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow Marzano causal teacher evaluation system in regard to deliberate practice. A Chi-square test was conducted in SPSS for Research Question 2. The independent variable in the test was whether or not a district followed the implementation model in regard to

41

deliberate practice as explained in the literature review. The dependent variable was

teacher performance ratings, as reported by the percentage of teachers in each of the three

categories (highly effective, effective, and needs improvement /developing/

unsatisfactory). The significance level for this test was $p = .05$; therefore, if the p value

for the test was below .05, the null hypothesis was rejected. The results from this test are

reported in Chapter 4 of this study.

Research Question 3

Research Question 3 sought to determine what relationship, if any, exists in

teacher performance ratings between school districts that followed and did not follow the

Marzano causal teacher evaluation system in regard to the number of formal observations

conducted. A Chi-square test was conducted in SPSS for research question three. The

independent variable in the test was whether or not a district followed the implementation

model in regard to the number of formal observations conducted as explained in the

literature review. The dependent variable was teacher performance ratings, as reported

by the percentage of teachers in each of the three categories (highly effective, effective,

and needs improvement/ developing/unsatisfactory). The significance level for this test

was $p = .05$; therefore, if the p value for the test was below .05, the null hypothesis was

rejected. The results from this test are reported in Chapter 4 of this study.

Research Question 4

Research Question 4 sought to determine what relationship, if any, exists in

teacher performance ratings between school districts that followed and did not follow the

Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations. A Chi-square test was conducted in SPSS for Research Question 4. The independent variable in the test was whether or not a district followed the implementation model in regard to conducting pre- and post-conferences during formal observations as explained in the literature review. The dependent variable was teacher performance ratings, as reported by the percentage of teachers in each of the three categories (highly effective, effective, and needs improvement/developing/unsatisfactory). The significance level for this test was $p = .05$; therefore, if the p value for the test was below .05, the null hypothesis was rejected. The results from this test are reported in Chapter 4 of this study.

Research Question 5

Research Question 5 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing. A Chi-square test was conducted in SPSS for Research Question 5. The independent variable in the test was whether or not a district followed the implementation model in regard to which elements were observed during formal observations as explained in the literature review. The dependent variable was teacher performance ratings, as reported by the percentage of teachers in each of the three categories (highly effective, effective, and needs improvement/developing/unsatisfactory). The significance level for this test was $p = .05$;

therefore, if the p value for the test was below .05, the null hypothesis was rejected. The results from this test are reported in Chapter 4 of this study.

Research Question 6

Research Question 6 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of informal observations. A Chi-square test was conducted in SPSS for Research Question 3. The independent variable in the test was whether or not a district followed the implementation model in regard to the number of informal observations conducted as explained in the literature review. The dependent variable was teacher performance ratings, as reported by the percentage of teachers in each of the three categories (highly effective, effective, and needs improvement/ developing/unsatisfactory). The significance level for this test was $p = .05$; therefore, if the p value for the test was below .05, the null hypothesis was rejected. The results from this test are reported in Chapter 4 of this study.

Research Question 7

Research Question 7 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs. A Chi-square test was conducted in SPSS for Research Question 7. The independent variable in the test was whether or not a district followed the implementation model in regard to which elements were observed during informal observations as

explained in the literature review. The dependent variable was teacher performance

ratings, as reported by the percentage of teachers in each of the three categories (highly

effective, effective, and needs improvement/ developing/unsatisfactory). The

significance level for this test was p = .05; therefore, if the p value for the test was below

.05, the null hypothesis was rejected. The results from this test are reported in Chapter 4

of this study.

Research Question 8

Research Question 8 sought to determine what relationship, if any, exists in

teacher performance ratings between school districts that followed and did not follow the

Marzano causal teacher evaluation system in regards to implementation of Domain 2,

Domain 3, and Domain 4 evaluations. A Chi-square test was conducted in SPSS for

Research Question 8. The independent variable in the test was whether or not a district

followed the implementation model in regard to observations in Domain 2 as explained in

the literature review. The dependent variable was teacher performance ratings, as

reported by the percentage of teachers in each of the three categories (highly effective,

effective, and needs improvement/ developing/unsatisfactory). The significance level for

this test was p = .05; therefore, if the p value for the test was below .05, the null

hypothesis was rejected. The results from this test are reported in Chapter 4 of this study.

Summary

This chapter restated the purpose of the research and the research questions. The

population for the research was the 67 public school districts in Florida, and the sample

was comprised of the 25 school districts using the Marzano causal teacher evaluation system during the 2012-2013 school year. All districts that used the Marzano causal teacher evaluation system reported implementation plans and reported teacher performance data were used. The implementation plans were read, and variables were classified using a data collection sheet. District teacher performance ratings were added to the data. Implementation variables were analyzed and compared to the research-based model explained in the literature review. All data were then imported into SPSS, where Chi-square tests were run for Research Questions 2-8 to determine what relationship, if any, existed between the implementation variables (independent variables) and teacher performance ratings (dependent variables). Results from the data analysis are presented in Chapter 4.

CHAPTER 4
RESULTS

Introduction

In this study, the researcher sought to understand the different ways in which the Marzano causal teacher evaluation system was being implemented in the state of Florida. Also investigated was how changes in implementation and implementation variables related to teacher performance ratings.

This chapter provides a detailed description of the results from this study. This chapter is divided into eight sections, one for each of the eight research questions. The section devoted to Research Question 1 is further subdivided to accommodate separate narrative descriptions of the 25 school districts and their implementation of the Marzano causal evaluation system during the 2012-2013 school year. The final section of this chapter contains a summary of the results of the study.

Research Question 1

Research Question 1 sought to identify the different ways in which 25 Florida school districts implemented the Marzano causal teacher evaluation system during the 2012-2013 school year. Following is a detailed description of the 25 school district evaluation implementation plans as related to the identified variables in this study.

District A

The district followed the research-based model related to the implementation of deliberate practice. Deliberate practice was calculated based on teacher improvement of

identified instructional elements, and this score was included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers with 1-3 years of experience were required to have a minimum of two formal observations during the school year. Teachers with four or more years of experience were required to have two formal observations during the school year. The number of required formal observations was below the recommended number of observations. No limits were placed on the number of formal observations that could be conducted.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations. However, during formal observations of first-year teachers the observer was required to establish design questions as a focus for the formal observations. The focus was to be collaboratively established between the observer and teacher. Each design question had multiple elements that could be observed, and multiple design questions were allowed to be a focus for the observation. Only elements within the selected design questions could be used for the observation. This did slightly limit the observed elements during the observation of first year teachers, but this alteration was very small and minimally limited the scope of elements observed with a small group of teachers.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance

evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with one to three years of experience were required to have a minimum of two informal observations. Teachers with four or more years of experience were required to have a minimum of one informal observation during the school year.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs. Teachers with one to three years of experience were required to have a minimum of four classroom walkthrough evaluations. Teachers with four or more years of experience were required to have a minimum of two classroom walkthrough evaluations during the school year. No limit was placed on the number of classroom walkthroughs that could be conducted.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1. The evaluation procedures guide identified specific sources of evidence that had to be used for the evaluation of these domains. For Domain 2, the sources were limited to: planning conference or

49

preconference, and artifacts. For Domain 3, sources of data were limited to: self-

assessment, reflection conference, conferences, discussions, and artifacts. For Domain 4

sources of data were limited to conferences, discussions, and artifacts. Although the

limiting of data sources was not indicated in the research, the sources of data appeared

broad enough to include almost all ways in which these domains could be observed.

District B

The district did not follow the research-based model related to the implementation

of deliberate practice. Deliberate practice was not included in teacher performance

ratings.

The district did not follow the research-based model related to the number of

formal observations that were to be conducted during the 2012-2013 school year. In

defining the number of formal observations to be conducted, the state-approved

documentation stated that a considerable amount must be used. This statement was not

quantified. Probationary contract teachers were required to have two formal observations

with no limit. All other teachers were required to have a minimum of one formal

observation, with no maximums. These minimums were lower than research-based

model.

The district followed the research-based model related to the scope of elements

observed during formal observations. Informal observations counted toward teacher

performance ratings. Elements observed were not limited, and no limit was placed on the

number of informal observations that could take place.

The district did not follow the research-based model related to pre- and post-conferencing during formal observations. Evaluators were not required to conduct pre-conferences with teachers. However, observers were required to conduct post conferences within 10 days of a formal observation.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward end-of-year teacher instructional practice scores, and the scope of observed elements were not limited during informal observations. In defining the number of informal observations to be conducted, the documentation stated that a considerable amount must be used. This statement was not quantified. A minimum of one informal observation was required of all employees, with no maximum.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district did not follow the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

There were two additional findings revealed during the review of the District B teacher evaluation system. It was discovered that during the 2013-2014 school year new elements were added to Domain 1. These elements were not part of the Marzano causal teacher evaluation system. Additionally, it was stated that a minimum of 45 data

markers, i.e., elements, were to be collected. There was nothing in the research base to suggest a minimum number of elements to be evaluated during a school year. Finally the district put minimum and maximum time limits on formal, informal, and classroom walkthrough evaluations. All of these changes appear to have occurred during the 2013-2014 school year, and did not impact this study.

District C

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was calculated based on teacher improvement of identified instructional elements, and this score was included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. First-year teachers were required to be formally observed a minimum of two times during the school year, with no maximum specified. Teachers with 1-3 years of service in the district were required to be formally observed once a year, with no maximum indicated. Teachers with 10 or more years of service in the district who were rated at effective or higher, were required to be observed once every three years. These minimums were below the recommended number of formal observations.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations. One finding in the review of the state-

approved district implementation plan was that the observation protocols used in the school district were altered. Though these look-fors assist in the observation process, they were removed from the protocol sheets.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were required to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings. Elements observed were not limited, and no limit was placed on the number of informal observations that could take place. District documentation was vague on the number of informal observations to be conducted. Additionally, there was confusion in the documentation in the difference between classroom walkthroughs and informal observations, i.e., the documentation refers to informal observations as classroom walkthroughs. The research base for the Marzano causal teacher evaluation system shows a difference in these two types of teacher observations. The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs did not count toward teacher performance ratings.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1. It appears these domains were only

observed at specific times, and all elements were evaluated when observing them. For the purpose of this study, the variables of implementing Domains 2, 3 and 4 were used.

District D

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Annual contract teachers were required to have two formal observations during a year. Continuing and professional service contract teachers were required to have one formal observation a year. This number of required observations was below the recommended amount of formal observations. Maximums were placed on the number of formal observations that could be conducted unless the administrator had a concern about employee performance.

The district did not follow the research-based model related to the scope of elements observed during formal observations. During formal observations, observers were limited to observing elements: 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 24, 25, 26, 27, 28,29, 30, 31, and 32.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance

evaluation system documentation stated that observers were required to conduct pre- and post-conferences during the formal observation process.

The district did not follow the research-based model related to the implementation of informal observations. Informal observations did not count toward teacher performance ratings.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs did not count toward teacher performance ratings.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings, but were weighted less than elements in Domain 1.

District E

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. The number of formal observations conducted followed the research-based model. Teachers with one to three years of experience were required to have a minimum of four formal observations, with no maximum. Teachers with four to nine years of experience were required to have a minimum of two formal observations, with no maximum. Teachers

with 10 or more years of experience were required to have a minimum of one formal observation, with no maximum. These were below the minimums identified in the research.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations. The scope of elements observed during formal observations followed the research-based model. During formal observations, all elements were observed. However, for first year teachers, only elements in specific design questions were used during formal observations. This is a slight, but minimal, change to the research-based model. Therefore, for the purpose of this study for the variable of observed formal elements, this variable was considered to be used correctly.

The district followed the research-based model related to pre-and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre-and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with one to three years of experience were required to have a minimum of five formal observations. Teachers with four to nine years of experience were required to have a minimum of two formal

observations. Teachers with 10 or more years of experience were required a minimum of one or two formal observations. These were required minimums, and no maximums were indicated.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs. Teachers with one to three years of experience were required to have a minimum of two classroom walkthroughs monthly. Teachers with four to nine years of experience were required to have a minimum of one classroom walkthrough monthly. Teachers with 10 or more years of experience were required a minimum of one classroom walkthrough monthly. These were required minimums, and no maximums were indicated.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

District F

No data were available for this school district.

District G

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district followed the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. For teachers working one to three years, a minimum of six formal observations were required. For teachers working four or more years, a minimum of three formal observations was required. There was no limit on the number of formal observations to be conducted.

The district did not follow the research-based model related to the scope of elements observed during formal observations. For the first three formal observations, observers were required to evaluate specific elements during each observation. The identified elements to observe changed for each of the formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. For all teachers in the district, a minimum of five informal observations was required.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs did not count toward teacher performance ratings.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings, but were weighted less than elements in Domain 1.

District H

The district followed the research-based model related to the implementation of deliberate practice. Deliberate practice was calculated based on teacher improvement of identified instructional elements, and this score was included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. For teachers working one to three years, a minimum of two formal observations was required. For teachers working four or more years, a minimum of one formal observation was required. No maximum was placed on the number of formal observations to be counted toward teacher performance ratings.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district did not follow the research-based model related to pre- and post-conferencing during formal observations. Teachers were required to submit a written lesson plan to the administrator who would observe, but no pre-conference was required. A post-conference was required. The district did state that one pre-conference would be

held at the beginning of the year between the principal and teacher to develop an instructional improvement plan for the school year. However, this type of conference was different from a pre-conference to discuss an upcoming formal observation.

The district did not follow the research-based model related to the implementation of informal observations. Evaluators were required to conduct 10 informal walkthroughs a year for teachers but were not required to give teachers feedback on these observations. Informal walkthroughs were defined as being teacher meetings, attending professional development, and other types of teacher actions. Informal observations were unannounced classroom observations.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1. These performance ratings were recorded as informal teacher observations.

District I

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Though all teachers in the district were required to have a minimum of two formal observations, there were no indicated maximums.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with less than four years of experience were required to have a minimum of six informal observations. Teachers with four or more years of experience were required to have a minimum of three informal observations.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during

classroom walkthroughs. All teachers in the district were required to have a minimum of three classroom walkthrough evaluations, but more were allowed.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

District J

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. First-year teachers were required to have three formal observations conducted; teachers who had taught for two or three years were required to have two formal observations conducted; and teachers who taught four or more years were required to have one formal observation conducted. These numbers were below the research-based recommended minimums. No limit was placed on the number of formal observations that could be conducted on teachers.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation required observers to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with less than three years of teaching experience were required to have a minimum of two informal observations. Teachers with three or more years of experience were not required to have any informal observations.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs. Teachers with less than three years of teaching experience were required to two classroom walkthroughs a month. Teachers with three or more years of experience were required to have one classroom walkthrough a month. These classroom walkthroughs counted toward teacher evaluations

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings, but were weighted less than elements in Domain 1.

District K

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers in their first year of teaching were required to have a minimum of two formal observations conducted, teachers teaching between two and nine years were required to have one formal observation conducted during the year, and teachers teaching 10 or more years were not required to have any formal observations conducted. These were below the recommended number of formal observations. No maximums were placed on the number of formal observations to be conducted.

The district did not follow the research-based model related to the scope of elements observed during formal observations. Limits were placed on the elements that could be observed during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of

64

informal observations that could take place. Teachers teaching from one to nine years were required to have one informal observation, and teachers teaching 10 or more years were not required to have any informal observations conducted.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs. Teachers teaching from one to nine years were required to have six classroom walkthroughs, and teachers teaching 10 or more years were not required to have any classroom walkthroughs.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings, but were weighted less than elements in Domain 1.

District L

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was used, but only counted in teacher performance ratings if it helped teacher performance ratings. A deliberate practice score should be included in teacher performance ratings regardless of whether it benefited or lowered a teacher's evaluation score.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. First year teachers were required to have two formal observations, and all other teachers were

required to have a minimum of one formal observation. The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district did not follow the research-based model related to pre- and post-conferencing during formal observations. The state- approved district performance evaluation system documentation stated that only formal observations on first year teachers required pre-conferences.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. First year teachers were required to have four informal observations, and all other teachers were required to have a minimum of two informal observations.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district did not follow the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 did not count toward teacher evaluations.

District M

The district followed the research-based model related to the implementation of deliberate practice. Deliberate practice was calculated based on teacher improvement of identified instructional elements, and this score was included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers with less than four years of experience were required to have two formal observations during the year, with no maximum. Teachers with four or more years of experience were required to have a minimum of one formal observation with no maximum.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with less than four years of

experience were required to have two informal observations during the year, with no maximum. Teachers with four or more years of experience were required to have a minimum of one informal observation with no maximum.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs. All teachers were required to have classroom walkthroughs, which count toward their evaluations, every nine weeks.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

District N

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers with less than four years of experience were required to have two formal observations. Teachers with four or more years of experience were required to have one formal observation.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation required that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with less than four years of experience were required to have six informal observations. Teachers with four or more years of experience were required to have four to six evaluations, depending on the scores of previous evaluations.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings, but were weighted less than elements in Domain 1.

District O

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that was to be conducted during the 2012-2013 school year. First year teachers were required to have two formal observations. All other teachers were required to have one formal observation. The district documentation also referred to unannounced formal observations. There was nothing in the research-based model about an unannounced formal observation. An unannounced formal observation would be, by definition, classified as an informal observation.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district did not follow the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that pre-conferences could be used, but that post conferences were required.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. First year teachers were required to have

70

four informal observations. All other teachers were required to have two informal observations.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings, but were weighted less than elements in Domain 1.

District P

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers with less than four years of experience were required to have two formal observations, with no maximum. Teachers with four or more years of experience were required to have one formal observation, with no maximum.

The district did not follow the research-based model related to the scope of elements observed during formal observations. The district limited formal observations to the following elements: 1-13 and 24-38.

71

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post conferences during the formal observation process

The district did not follow the research-based model related to the implementation of informal observations. Teachers with less than four years of experience were required to have four informal observations. Teachers with four or more years of experience were required to have two informal observations. For all teachers, the first informal observation of the year did not count toward their evaluations. For all teachers, scores for informal observations conducted after the required minimum would replace previous performance ratings if the new score was higher than the previous scores.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

District Q

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers with less than three years of experience were required to have between two and four formal observations. No more than four formal observations were allowed to count towards teacher evaluation ratings. Teachers with three or more years of experience were required to have between one and two formal observations. No more than two formal observations were allowed to count toward teacher evaluation ratings

The district did not follow the research-based model related to the scope of elements observed during formal observations. The scope of elements to be observed was limited during formal observations. The elements to be observed were limited to one or two elements selected by the teacher at the beginning of the year. One additional school wide focus element could also be evaluated. No other elements were allowed to be evaluated during the observation.

The district did not follow the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district did not follow the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and a limit was placed on the number of informal observations that could take place. Teachers with less than three years of experience were required to have between two and four informal observations. No more than four informal observations were allowed to count toward teacher evaluation ratings. Teachers with three or more years of experience were required to have between one and two informal observations. No more than two informal observations were allowed to count toward teacher evaluation ratings

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

District R

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers were required to have one formal observation; the option of either one classroom walkthrough, one informal observation or one formal observation would be included in teachers' performance ratings.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district did not follow the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district did not follow the research-based model related to the implementation of informal observations. A maximum of only one informal observation could be used in a teacher's end-of-year evaluation.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs. All teachers were required to have at least one classroom walkthrough count toward their end of the year evaluation.

The district did not follow the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

The district followed the research-based model related to the implementation of deliberate practice. Deliberate practice was calculated based on teacher improvement of identified instructional elements, and this score was included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations required to be conducted during the 2012-2013 school year. First year teachers were required to have a minimum of three formal observations conducted during the year. All other teachers were required to have one formal observation conducted during the year.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations did not count toward teacher performance ratings.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs did not count toward teacher performance ratings.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 did not count toward teacher performance ratings.

District T

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings. The district followed the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers with less than four years of experience were required to have four formal observations. Teachers with four or more years of experience were required to have two formal observations. There was no limit to the number of formal observations that counted toward teacher performance ratings.

The district followed the research-based model that related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

77

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with less than four years of experience were required to have four informal observations. Teachers with four or more years of experience were required to have two informal observations. There was no limit to the number of informal observations that counted toward teacher performance ratings.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

District U

The Florida Department of Education reported that this district implemented the Marzano casual evaluation system. However, the school district rewrote the observation

protocols for the Marzano casual evaluation system. This means that during classroom observations, teachers were evaluated based on different criteria from other counties using the same system. This complete rewrite was such a significant change to the evaluation system that the data from this district were excluded from the study.

<center>*District V*</center>

No data were available for this school district.

<center>*District W*</center>

The district followed the research-based model related to the implementation of deliberate practice. Deliberate practice was not included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year.

Teachers with less than three years of teaching experience were required to have two formal observations, with no maximum. Teachers with three or more years of experience were required to have one formal observation, with no maximum.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with less than three years of teaching experience were required to have three informal observations, with no maximum. Teachers with three or more years of experience were required to have two informal observations, with no maximum.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs did not count toward teacher performance ratings.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings but were weighted less than elements in Domain 1.

District X

The district did not follow the research-based model related to the implementation of deliberate practice. Deliberate practice was calculated based on teacher improvement of identified instructional elements, and this score was included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers in their first year of teaching in the school district were required to have two

formal observations, with no maximum. All other teachers were required to have one formal observation, with no maximum.

The district followed the research-based model related to the scope of elements observed during formal observations. The district did not limit the observable instructional elements during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place.

The district did not follow the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs did not count toward teacher performance ratings.

The district did not follow the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 did not count toward teacher performance ratings.

District Y

The district followed the research-based model related to the implementation of deliberate practice. Deliberate practice was calculated based on teacher improvement of identified instructional elements and this score was included in teacher performance ratings.

The district did not follow the research-based model related to the number of formal observations that were to be conducted during the 2012-2013 school year. Teachers with less than three years of experience were required to have two formal observations, with no maximum. Teachers with three or more years of experience were required to have one formal observation.

The district did not follow the research-based model related to the scope of elements observed during formal observations. The evaluator was required to evaluate a specific design question during formal observations.

The district followed the research-based model related to pre- and post-conferencing during formal observations. The state-approved district performance evaluation system documentation stated that observers were to conduct pre- and post-conferences during the formal observation process.

The district followed the research-based model related to the implementation of informal observations. Informal observations counted toward teacher performance ratings, elements observed were not limited, and no limit was placed on the number of informal observations that could take place. Teachers with less than three years of experience were required to have two informal observations, with no maximum.

Teachers with three or more years of experience were required to have one informal observation.

The district followed the research-based model related to the implementation of classroom walkthroughs. Classroom walkthroughs counted toward teacher performance ratings, and no limit was placed on the number of elements to be observed during classroom walkthroughs. All teachers were required to have two classroom walkthroughs, which counted toward teacher performance ratings. No maximum was placed on the number of classroom walkthroughs that could be counted toward teacher performance ratings.

The district followed the research-based model related to the implementation of Domains 2, 3, and 4. Domains 2, 3, and 4 counted toward teacher performance ratings, but were weighted less than elements in Domain 1.

<div align="center">Research Question 2</div>

Research Question 2 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to deliberate practice. The null hypothesis was that no relationship exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to deliberate practice.

To research this question, teacher performance ratings between school districts that followed and did not follow the deliberate practice implementation model were

compared using a Chi-square analysis. Table 2 shows the results of the Chi-square test

for deliberate practice. Based on the chi square test for independence, the null hypothesis

was rejected, and a statistically significant relationship was found between districts that

implemented deliberate practice and districts that did not implement deliberate practice

according to the research model (p <.00).

Table 2

Chi-square Test for Deliberate Practice

Descriptor	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	72.432	2	.000
N of Valid Cases	2101		

Table 3 shows a comparison of teacher performance ratings for deliberate

practice. In districts that implemented deliberate practice according to the research

model, 41% of teachers were rated as highly effective, 55% were rated effective, and 5%

were rated needs improvement/developing/unsatisfactory. In districts that did not

implement deliberate practice according to the research model, 23% of teachers were

rated as highly effective, 73% were rated effective, and 4% were rated needs

improvement/developing/unsatisfactory. In districts that implemented deliberate practice

according to the research model, more teachers were rated highly effective (41%

compared to 23%), and fewer were rated effective (55% compared to 73%). In districts

that implemented deliberate practice, about the same number of teachers were rated as

improvement/developing/unsatisfactory as in school districts that did not implement

deliberate practice (5% compared to 4%).

Table 3

Comparison of Teacher Performance Ratings for Deliberate Practice

| | Implementation Model | |
Teacher Performance Category	Followed	Did Not Follow
Highly Effective	41%	23%
Effective	55%	73%
Improving/Developing/Unsatisfactory	5%	4%

Research Question 3

Research Question 3 sought to determine what relationship, if any, exists in

teacher performance ratings between school districts that followed and did not follow the

Marzano causal teacher evaluation system in regard to the number of formal observations

conducted. The null hypothesis was that no relationship exists in teacher performance

ratings between school districts that followed and did not follow the Marzano causal

teacher evaluation system in regard to the number of formal observations conducted. To

research this question, teacher performance ratings between school districts that followed

and did not follow the model in regard to number of formal observations conducted were

compared using a Chi-square analysis. Table 4 shows the results of this analysis. Based

on the Chi-square test for independence, the null hypothesis was supported, and no

significant relationship was found between districts that followed and those that did not follow the model in regard to number of formal observations conducted (p = .25). The adherence to the recommended number of formal observations conducted appeared to have no statistically significant relationship with teacher evaluation ratings in school districts that did or did not follow the implementation model.

Table 4

Chi-square Test for Number of Formal Observations

Descriptor	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.741	2	.254
N of Valid Cases	2101		

Table 5 shows a comparison of teacher performance ratings for number of formal observations. In districts that followed the research model in regard to the number of formal observations conducted, 24% of teachers were rated as highly effective, 73% were rated effective, and 3% were rated as needs improvement/developing/ unsatisfactory. In districts that did not follow the research model in regard to the number of formal observations conducted, 28% of teachers were rated as highly effective, 67% were rated effective, and 4% were rated needs improvement/developing/unsatisfactory.

Table 5

Comparison of Teacher Performance Ratings for Number of Formal Observations

	Implementation Model	
Teacher Performance Category	Followed	Did Not Follow
Highly Effective	24%	28%
Effective	73%	67%
Improving/Developing/Unsatisfactory	3%	4%

Research Question 4

Research Question 4 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations. The null hypothesis was that no relationship exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations.

To research this question, teacher performance ratings between school districts that followed and did not follow the model in regard to the scope of elements observed in formal observations were compared using a Chi-square analysis. Table 6 shows the results of this analysis. Based on the Chi-square test for independence, the null hypothesis was rejected and a significant relationship was found between districts that

followed and did not follow the implementation model in regard to the scope of elements observed in formal observations (p <.001).

Table 6

Chi-square Test for Scope of Elements Observed in Formal Observations

Descriptor	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.381	2	.006
N of Valid Cases	2101		

Table 7 shows a comparison of teacher performance ratings for scope of elements observed in formal observations. In districts that followed the research model in regard to the scope of elements observed in formal observations, 26% of teachers were rated as highly effective, 69% were rated effective, and 5% were rated needs improvement/developing/ unsatisfactory. In districts that did not follow the research model in regard to the scope of elements observed in formal observations, 32% of teachers were rated as highly effective, 66% were rated effective, and 3% were rated needs improvement/developing/unsatisfactory. In districts that followed the research model in regard to the scope of elements observed in formal observation, fewer teachers were rated as highly effective (26% compared to 32% highly effective). In districts that followed the research model in regard to the scope of elements, more teachers were rated as improvement/developing/unsatisfactory as in the school districts that did not follow the implementation model (5% compared to 3%).

Table 7

Comparison of Teacher Performance Ratings for Scope of Elements Observed in Formal Observations

| | Implementation Model | |
Teacher Performance Category	Followed	Did Not Follow
Highly Effective	26%	32%
Effective	69%	66%
Improving/Developing/Unsatisfactory	5%	3%

Research Question 5

    Research Question 5 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing. The null hypothesis was that no relationship exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing.

    To research this question, teacher performance ratings between school districts that followed and did not follow the model in regard to the implementation of pre- and post-conferencing were compared using a Chi-square analysis. Table 8 shows the results of this analysis. Based on the Chi-square test for independence, the null hypothesis was rejected, and a significant relationship was found between districts that followed and did

not follow the implementation model in regard to the implementation of pre- and post-conferencing (p <.001).

Table 8

Chi-square Test for Pre- and Post-conferencing Implementation

Descriptor	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	19.081	2	.000
N of Valid Cases	2101		

Table 9 shows a comparison of teacher performance ratings for pre- and post-conferencing implementation. In districts that followed the research model in regard to the implementation of pre- and post-conferencing, 27% of teachers were rated as highly effective, 68% were rated effective, and 5% were rated needs improvement/developing/unsatisfactory. In districts that did not follow the research model in regard to the implementation of pre- and post-conferencing, 30% of teachers were rated as highly effective, 69% were rated effective, and 1% were rated needs improvement/developing/ unsatisfactory. In districts that followed the research model in regard to the implementation of pre- and post-conferencing, more teachers were rated as needs improvement/developing/unsatisfactory than in districts that did not follow the implementation model (5% compared to 1%).

Table 9

Comparison of Teacher Performance Ratings for Pre- and Post-conferencing Implementation

	Implementation Model	
Teacher Performance Category	Followed	Did Not Follow
Highly Effective	27%	30%
Effective	68%	69%
Improving/Developing/Unsatisfactory	5%	1%

Research Question 6

Research Question 6 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of informal observations. The null hypothesis was that no relationship exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of informal observations.

To research this question, teacher performance ratings between school districts that followed and did not follow the model in regard to the implementation of informal observations were compared using a Chi-square analysis. Table 10 shows the results from this analysis. Based on the Chi-square test for independence, the null hypothesis was rejected, and a significant relationship was found between districts that followed and

those that did not follow the implementation model in regard to the implementation of informal observations (p <.001).

Table 10

Chi-square Test for Implementation of Informal Observations

Descriptor	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	17.675	2	.000
N of Valid Cases	2101		

Table 11 shows a comparison of teacher performance ratings for Implementation of informal observations. In districts that followed the research model in regard to the implementation of informal observations, 27% of teachers were rated as highly effective, 68% were rated effective, and 5% were rated needs improvement/developing/unsatisfactory. In districts that did not follow the research model in regard to the implementation of informal observations, 32% of teachers were rated as highly effective, 67% were rated effective, and 1% were rated needs improvement/developing/unsatisfactory. In districts that followed the research model in regard to the implementation of informal observations, more teachers were rated as needs improvement/developing/unsatisfactory (5% compared to 1%) and fewer teachers were rated as highly effective, than in school districts that did not follow the implementation model (27% compared to 32%).

Table 11

Comparison of Teacher Performance Ratings for Implementation of Informal Observations

| | Implementation Model | |
Teacher Performance Category	Followed	Did Not Follow
Highly Effective	27%	32%
Effective	68%	67%
Improving/Developing/Unsatisfactory	5%	1%

Research Question 7

Research Question 7 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs. The null hypothesis was that no relationship exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs.

To research this question, teacher performance ratings between school districts that followed and did not follow the model in regard to the implementation of classroom walkthroughs were compared using a Chi-square analysis. Table 12 shows the results of this analysis. Based on the Chi-square test for independence, the null hypothesis was rejected, and a significant relationship was found between districts that followed and did

not follow the implementation model in regard to the implementation of formal

observations (p <.001).

Table 12

Chi-square Test for Implementation of Classroom Walkthroughs

Descriptor	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	39.213	2	.000
N of Valid Cases	2101		

Table 13 shows a comparison of teacher performance ratings for Implementation

of classroom walkthroughs. In districts that followed the research model in regard to the

implementation of classroom walkthroughs, 34% of teachers were rated as highly

effective, 63% were rated effective, and 3% were rated needs

improvement/developing/unsatisfactory. In districts that did not follow the research

model in regard to the implementation of classroom walkthroughs, 22% of teachers were

rated as highly effective, 73% were rated effective, and 5% were rated as needs

improvement/developing/unsatisfactory. In districts that followed the research model in

regard to the implementation of classroom walkthroughs, more teachers were rated as

highly effective (34% compared to 22%) and fewer teachers were rated as effective than

in school districts that did not follow the implementation model (63% compared to 73%).

Table 13

Comparison of Teacher Performance Ratings for Implementation of Classroom Walkthroughs

	Implementation Model	
Teacher Performance Category	Followed	Did Not Follow
Highly Effective	34%	22%
Effective	63%	73%
Improving/Developing/Unsatisfactory	3%	5%

Research Question 8

Research Question 8 sought to determine what relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to implementation of Domain 2, Domain 3, and Domain 4 evaluations. The null hypothesis was that no relationship exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of Domain 2, Domain 3, and Domain 4 evaluations.

To research this question, teacher performance ratings between school districts that followed and did not follow the model in regard to the implementation of Domain 2, Domain 3, and Domain 4 evaluations were compared using a Chi-square analysis. Table 14 shows the results of this analysis. Based on the Chi-square test for independence, the null hypothesis was rejected, and a significant relationship was found in teacher

performance ratings between districts that followed and did not follow the

implementation model in regard to the implementation of Domain 2, Domain 3, and

Domain 4 evaluations (p <.001).

Table 14

Chi-square Test for Implementation of Evaluations: Domains 2, 3, and 4

Descriptor	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	75.477	2	.000
N of Valid Cases	2101		

Table 15 shows a comparison of teacher performance ratings for Implementation

of Domain 2, Domain 3, and Domain 4 evaluations. In districts that followed the

research model in regard to the implementation of Domain 2, Domain 3, and Domain 4

evaluations, 31% of teachers were rated as highly effective, 66% were rated effective,

and 3% were rated needs improvement/developing/unsatisfactory. In districts that did not

follow the research model in regard to the implementation of Domain 2, Domain 3, and

Domain 4 evaluations, 14% of teachers were rated as highly effective, 77% were rated

effective, and 9% were rated needs improvement/developing/unsatisfactory. In districts

that followed the research model more teachers were rated as highly effective (31%

compared to 14%) and fewer teachers were rated as effective, than in districts that did not

follow the implementation model (66% compared to 77%). Also, in districts that

followed the research model in regard to the implementation of Domain 2, Domain 3, and

Domain 4 evaluations, fewer teachers were rated needs improvement/developing/

unsatisfactory , than in districts that did not follow the implementation model (3%

compared to 9%).

Table 15

Comparison of Teacher Performance Ratings for Implementation of Evaluations:
Domains 2, 3, and 4

| | Implementation Model | |
Teacher Performance Category	Followed	Did Not Follow
Highly Effective	31%	14%
Effective	66%	77%
Improving/Developing/Unsatisfactory	3%	9%

<u>Summary</u>

Table 16 presents a summary of how the Marzano causal teacher evaluation

system was implemented in the state of Florida during the 2012-2013 school year. A "y"

indicates that the implementation variable was followed according to the research-based

model. An "n" indicates that the implementation variable was not followed according to

the research-based model.

Table 16

Summary of School District Implementation Variables

				Implementation Variables			
School District	Deliberate Practice	Number of Formals	Scope of Formal Elements	Pre & Post Conferences	Informal Observations	Classroom Walkthroughs	Domains 2,3, 4
A	Y	N	Y	Y	Y	Y	Y
B	N	N	Y	N	Y	N	N
C	N	N	Y	Y	Y	N	Y
D	N	N	N	Y	N	N	Y
E	N	N	Y	Y	Y	Y	Y
F				data not available			
G	N	Y	N	Y	Y	N	Y
H	Y	N	Y	N	N	Y	Y
I	N	N	Y	Y	Y	Y	Y
J	N	N	Y	Y	Y	Y	Y
K	N	N	N	Y	Y	Y	Y
L	N	N	Y	N	Y	N	N

School District	Deliberate Practice	Number of Formals	Scope of Formal Elements	Pre & Post Conferences	Informal Observations	Classroom Walkthroughs	Domains 2, 3, 4
M	Y	N	Y	Y	Y	Y	Y
N	N	N	Y	Y	Y	N	Y
O	N	N	Y	N	Y	Y	Y
P	N	N	N	Y	N	N	Y
Q	N	N	N	N	N	N	Y
R	N	N	Y	N	N	Y	N
S	Y	N	Y	Y	Y	Y	Y
T	N	Y	Y	Y	Y	Y	Y
U				data not available			
V				data not available			
W	Y	N	Y	Y	Y	N	Y
X	N	N	Y	Y	Y	N	N
Y	Y	N	N	Y	Y	Y	Y

The data shown in Table 16 indicates that the implementation of the Marzano causal teacher evaluation system varied widely among the school districts in the state of Florida during the 2012-2013 school year. No school district implemented all of the implementation variables according to the researched-based model. The variable most widely implemented incorrectly was the number of formal observations to be conducted (implementation variable two). Only two districts followed the implementation model according to the number of formal observations to be conducted. The variable most widely implemented correctly, according to the research-based model, was the implementation of Domain 2, Domain 3, and Domain 4 evaluations (implementation variable seven). Only four of the school districts did not implement this variable according to the research model.

Table 17 presents a summary of the statistical analyses conducted for Research Questions 2-8. As evidenced in the results of the Chi-square analyses, with the exception of only one variable, there was a relationship between teacher performance ratings and whether or not a district implemented a particular variable as it was intended to be implemented in the model.

Table 17

Summary of Statistical Significance of Implementation Variables

Implementation Variables	Statistically Significant Relationship
Deliberate practice	Yes
Number of formal observations	No
Scope of formal elements	Yes
Pre- and post-conferencing	Yes
Informal observations	Yes
Classroom walkthroughs	Yes
Domains 2, 3, and 4	Yes

CHAPTER 5 DISCUSSION

Introduction

The results of the research study were presented in Chapter 4. This chapter contains a summary of the study findings and a discussion of how those findings may be applied in practice. In this chapter, a brief summary of the study reviews the problem of the study, the purpose of the study, the research questions, the theoretical framework, the methodology and the findings of the study. The second section of the chapter contains a discussion of the findings and recommendations and is followed by recommendations for further research in which suggestions are offered as to how this research topic could be extended. The final section of this chapter provides a summative statement about the research study.

Summary of the Study

During the 2012-2013 school year, 25 school districts in the state of Florida implemented the Marzano causal evaluation system to evaluate teacher performance. Implementation of this system varied widely among school districts. At the time of the present study, no research had been conducted to investigate the different ways in which the Marzano causal teacher evaluation system had been implemented. Additionally, no research had been conducted to determine how variations in implementation may relate to teacher performance ratings.

The purpose of this study was to analyze the different ways in which Florida school districts implemented the Marzano causal teacher evaluation system and to see

what relationships may exist between different implementation variables and teacher performance ratings.

This study had eight research questions. Those questions were:

1. What are the different ways in which 25 Florida school districts implemented the Marzano causal teacher evaluation system during the 2012-2013 school year?

2. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to deliberate practice?

3. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the number of formal observations conducted?

4. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations?

5. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing?

6. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher

evaluation system in regard to the implementation of information observations?

7. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs?

8. What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to implementation of Domain 2, Domain 3, and Domain 4 evaluations?

The theoretical framework on which this study was based was Taylor's theory of scientific management.

The study included all school districts in the state of Florida who used the Marzano causal teacher evaluation system to evaluate teachers during the 2012-2013 school year. Data for the study were collected from the Florida Department of Education's website on which all districts have been required to report implementation guides and teacher performance ratings. District implementation plans were read and compared to the research-based model, and a narrative description of district implementation was written for each of the school districts to answer Research Question 1.

District implementation, related to the seven implementation variables, was recorded on a data collection sheet. Teacher performance ratings for each district were

also added to the data collection sheet. The results of the data analysis conducted to respond to Research Question 1 were summarized on the data collection sheet (Appendix C).

For Research Questions 2-8, teacher performance ratings for districts following the implementation variable were compared to those of teachers in districts not following the implementation variables. Using these data, a Chi-square test for relationship was conducted for each of the seven implementation variables.

There was a statistically significant relationship between teacher performance ratings among districts that implemented deliberate practice according to the research model and districts that did not implement deliberate practice according to the research based model.

There was no statistically significant relationship between teacher performance ratings among districts that followed the research model in relation to the number of formal observations conducted on teachers and districts that did not follow the research model in relation to the number of formal observations conducted.

There was a statistically significant relationship between teacher performance ratings among districts that followed the research model in relation to the scope of elements observed during formal observations and districts that did not follow the research model in relation to the scope of elements observed during formal observations.

There was a statistically significant relationship between teacher performance ratings among districts that followed the research model in relation to pre- and post-

105

conferencing during formal observations and districts that did not follow the research model in relation to pre- and post-conferencing during formal observations

There was a statistically significant relationship between teacher performance ratings among districts that followed the research model in relation to the implementation of informal observations and districts that did not follow the research model in relation to the implementation of informal observations.

There was a statistically significant relationship between teacher performance ratings among districts that followed the research model in relation to the inclusion of classroom walkthrough observations on teacher performance ratings and districts that did not follow the research model in relation to the inclusion of classroom walkthrough observations on teacher performance ratings.

There was a statistically significant relationship between teacher performance ratings among districts that followed the research model in relation to the inclusion of Domains 2, 3, and 4 on teacher performance ratings and districts that did not follow the research model in relation to the inclusion of Domains 2, 3, and 4 on teacher performance ratings.

Discussion of the Findings and Recommendations

Research Question 1

What are the different ways in which 25 Florida school districts implemented the Marzano causal teacher evaluation system during the 2012-2013 school year?

The findings from this research study showed that implementation of the Marzano causal teacher evaluation system in the state of Florida varied widely. These variations deviated from how the system was intended to be used. No districts analyzed had fully implemented the Marzano causal evaluation system as intended. Furthermore, the changes in system implementation often did not follow research-based best practice. The analysis and review of district implementation plans showed a lack of understanding of the Marzano causal evaluation system and how it was intended to be implemented. A lack of common language and understanding of common terms, e.g., informal observations and classroom walkthroughs, further supported this observation.

Based on the findings from this study it is suggested that educational leaders implementing the Marzano causal evaluation system review and improve their understanding of how the system was intended to be implemented. An understanding of common terms such as informal observations and classroom walkthroughs should be clarified by reviewing the available literature and research on the Marzano causal evaluation system, teacher evaluation systems in general, and the educational supervision process.

Research Question 2

What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to deliberate practice?

The findings from this study showed a relationship in teacher instructional practice scores between districts that did and did not implement the deliberate practice component of the Marzano causal teacher evaluation system. Not implementing deliberate practice according to the research-based model may result in inaccurate teacher performance ratings. The recommendation from this finding is that deliberate practice should be implemented according to the research model to achieve the most accurate teacher performance ratings.

Research Question 3

What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the number of formal observations conducted?

The findings from this study showed no relationship in teacher instructional practice scores between districts that did and did not implement the appropriate number of formal observations as outlined in the research model. Formal observations are part of the clinical supervision process. Researchers (Goldhammer, 1969; Marzano, 2011) have shown that the clinical supervision process supports teachers by providing a coaching and feedback model for teachers to learn. However, the results of this study suggested that the specific number of formal observations was not related to final teacher performance ratings. Because there is very specific criteria in the research-based model as to how many formal observations should be conducted for teachers in a given year, it is recommended that educators adhere to the criteria.

Research Question 4

What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the scope of elements observed in formal observations?

The findings from this study showed a relationship in teacher instructional practice scores between districts that did and did not limit the scope of elements observed during formal observations. This indicated that by limiting or altering the elements observed during formal observations, teacher performance ratings could be inaccurate. The recommendation, based on this finding, is that the number of elements observed should not be limited or altered during formal observations.

Research Question 5

What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of pre- and post-conferencing?

The findings from this study showed a relationship in teacher instructional practice scores between districts that did and did not implement pre- and post-conferencing during formal observations. This indicated that failure to implement pre- and post-conferencing during formal observations could result in inaccurate teacher performance ratings. The recommendation, based on this finding, is that pre- and post-conferencing should be conducted during the formal observation process.

Research Question 6

What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of informal observations?

The findings from this study showed a relationship in teacher instructional practice scores between districts that did and did not include informal observations in teacher performance ratings. This revealed that not including informal observations in teacher performance ratings could lead to inaccurate teacher performance ratings. The recommendation, based on this finding, was that informal observations be implemented according to the research-based model and included in teacher evaluations.

Research Question 7

What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow the Marzano causal teacher evaluation system in regard to the implementation of classroom walkthroughs?

The findings from this study showed a relationship in teacher instructional practice scores between districts that did and did not implement classroom walkthroughs as a part of the Marzano causal teacher evaluation system. This shows that by not including classroom walkthroughs in teacher performance ratings, teacher performance ratings could be inaccurate. The recommendation from this finding is that classroom walkthroughs be implemented according to the research-based model and included in teacher evaluations.

Research Question 8

What relationship, if any, exists in teacher performance ratings between school districts that followed and did not follow Marzano causal teacher evaluation system in regard to implementation of Domain 2, Domain 3, and Domain 4 evaluations?

The findings from this study showed a relationship in teacher instructional practice scores between districts that did and did not include Domains 2, 3, and 4 evaluations in teacher performance ratings. Thus, not including Domains 2, 3, and 4 evaluations in teacher performance ratings could result in inaccurate teacher performance ratings. The recommendation from this finding, therefore, is that Domain 2, Domain 3, and Domain 4 evaluations be implemented according to the research model and included in teacher evaluations.

Recommendations for Further Research

As reported in the literature review, there were no validity and reliability studies found on the Marzano causal teacher evaluation system, as a teacher evaluation tool. It is, therefore, recommended that such studies be initiated.

The goal of the study was to analyze the different ways in which the Marzano causal evaluation system has been implemented in the state of Florida and to determine what relationships, if any, exist in teacher evaluation performance ratings between districts that did and did not follow the research-based model. As districts continue to improve on the ways in which they are implementing the Marzano causal teacher evaluation system, data about implementation strategies should continue to be gathered.

As new variables become apparent, they should be tested to investigate relationships of teacher performance ratings in districts implementing the system differently. As research in this area grows, it may be possible to establish causal relationships and strengthen the research base, showing how the Marzano causal evaluation system should be implemented.

One stated purpose of the Marzano causal evaluation system has been to provide improvement feedback to teachers on their instructional practice. Further research should be conducted to see how effective the feedback provided through the Marzano causal evaluation system is to teacher improvement. How changes in implementation affect feedback given to teachers should also be studied.

Summary

The findings of this study show that the Marzano causal teacher evaluation system was not implemented correctly during the 2012-2013 school year in the state of Florida and that by not implementing the system correctly teacher performance ratings may be inaccurate. In seven of eight implementation variables, there were relationships between teacher performance ratings and changes in implementation from the research-based model. The analysis of the data supported the implementation of the Marzano causal evaluation system according to the research-based model and that changing the variables may erode the accuracy of teacher performance ratings.

The Marzano causal teacher evaluation system has been designed to provide teachers with ongoing performance feedback to allow them the opportunity to improve

their instructional techniques. If variables in implementation are altered, teacher performance data can be inaccurate, and, in turn, teacher feedback may also be inaccurate. Thus, incorrect implementation may not contribute to the full extent in improving teachers through the clinical supervision process.

The review of the district teacher evaluation plans shows that there has been confusion about the Marzano causal teacher evaluation system. For example, there are different definitions among school districts for basic terms such as classroom walkthroughs and informal observations. During the 2012-2013 school year, no school district implemented the Marzano causal teacher evaluation system exactly as it was designed to be implemented. In some cases, as in the case of pre- and post-conferencing, the system was implemented contrary to what researchers have advocated as best practice. This misuse and misunderstanding cannot be dismissed when analyzing the effectiveness of the Marzano system.

As future researchers begin to analyze the effectiveness of the Marzano causal evaluation system, attention must also be given to implementation mode. The system cannot be blamed if it is not used correctly. An example of this can be seen in the number of struggling teachers identified using the new evaluation system during the 2012-2013 school year. One of the intended and stated purposes of this new system was to reduce teacher evaluation rating inflation. With fewer than 1% of teachers being identified as struggling with the new system, some might state the new system did not meet this goal. However, the system may not be to blame. Improper implementation may be the problem. This presents an additional hypothesis. One could question whether

previous generation evaluation systems such as the FPMS were wrongly blamed for inflated teacher performance ratings. Perhaps previous generation teacher evaluations were not implemented according to best practice. The problem with teacher evaluation systems may not be the tools used but the various ways in which the systems are interpreted and implemented.

This study was intended to provide educational leaders and policymakers with support to ensure the proper implementation of the Marzano causal teacher evaluation system. There is clearly a proper way to implement the system. This requires being attentive to implementation variables so as not to erode the accuracy of teacher performance ratings. When purchasing expensive, new teacher evaluation systems, districts should be dedicated to implementing these systems as intended for the desired outcomes.

APPENDIX A
UCF INSTITUTIONAL REVIEW BOARD APPROVAL

University of
Central Florida

NOT HUMAN RESEARCH DETERMINATION

From :   **UCF Institutional Review Board #1**
          FWA00000351, IRB00001138

To   :   **Matthew Phillips**

Date :   **September 24, 2013**

Dear Researcher:

On 9/24/2013 the IRB determined that the following proposed activity is not human research as defined by DHHS regulations at 45 CFR 46 or FDA regulations at 21 CFR 50/56:

Type of Review:	Not Human Research Determination
Project Title:	AN ANALYSIS OF THE VARIABLES IN IMPLEMENTATION OF THE MARZANO CAUSAL TEACHER EVALUATION SYSTEM IN THE STATE OF FLORIDA
Investigator:	Matthew Phillips
IRB ID:	SBE-13-09597
Funding Agency:	
Grant Title:	
Research ID:	N/A

University of Central Florida IRB review and approval is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are to be made and there are questions about whether these activities are research involving human subjects, please contact the IRB office to discuss the proposed changes.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 09/24/2013 02:59:59 PM EDT

Joanne Muratori

IRB Coordinator

APPENDIX B
FLORIDA PERFORMANCE MEASUREMENT SYSTEM (FPMS)
OBSERVATION FORM

FLORIDA DEPARTMENT OF EDUCATION DIVISION OF HUMAN RESOURCE DEVELOPMENT □ **THE SCHOOL DISTRICT OF OSCEOLA COUNTY, FLORIDA**
FLORIDA PERFORMANCE MEASUREMENT SYSTEM SCREENING/SUMMATIVE OBSERVATION INSTRUMENT

No. of Students Not Engaged   1 ☐   2 ☐   3 ☐   4 ☐

FRAME FACTOR INFORMATION (PLEASE PRINT)

Teacher's Name

List, First, Middle

ID# _____

Institution of Graduation: _____ Inst#_____

Graduated from a College of Education   ☐ 1.YES   ☐ 2. NO

Number of Complete Years of teaching Experience: _____

District Name: _____ No._____

School Name: _____ No._____

Observer's Name

List, First, Middle

ID# _____

Position: ☐ 1.Principal ☐ 2.Asst. Principal ☐ 3.Teacher ☐ 4.Other

Class: _____ Grade Level: _____

Subject Area Observed

☐ 1. Language Arts   ☐ 9. Home Economics
☐ 2. Foreign Language   ☐ 10. Other Vocational Ed.
☐ 3. Social Sciences   ☐ 11. Arts
☐ 4. Mathematics   ☐ 12. Music
☐ 5. Science   ☐ 13. Exceptional Stud. Ed
☐ 6. Physical Ed., ROTC   ☐ 14. Other (specify)
☐ 7. Business Ed., DCT,CBE _____
☐ 8. Industrial Arts/Education

Type of Classroom/Facility in Which the Observation Occured
☐ 1. Regular Classroom – Self Contained, Open, Pod
☐ 2. Laboratory or Shop
☐ 3. Field, Court, Gymnasium
☐ 4. Media Room, Library, or

Total Number of Students in Class: _____

Observation Information
Date: _____
Type of Observation
☐ 1. New Teacher ☐ 2. Dist. Assess. ☐ 3. Other (specify)_____

Screening Obs.   ☐ 1.   ☐ 2.   ☐ 3.   ☐ 4.
Summative Obs.   ☐ 1.   ☐ 2.   ☐ 3.   ☐ 4.

Time Observation Begins: _____ Ends_____

Test Begins: _____ Ends_____

Methods Used in the Observed Lesson

☐ 1. Lecture   ☐ 2. Interaction/Discussion
☐ 3. Independent Study/Lab or Shop Work

Teacher's Signature: _____
(The signature of the teacher does not necessarily imply agreement with this observation, but rather, acknowledges that it has been discussed with the observer.)

Observer's Signature:_____

DOMAIN			TOT. FREQ	FREQUENCY	FREQUENCY	TOT. FREQ	
3.0 INSTRUCTIONAL ORGANIZATION AND DEVELOPMENT	1. Begins instruction promptly						1. Delays
	2. Handles Materials in and orderly manner						2. Does not organize material systematically
	3. Orients students to classwork/maintains academic focus						3. Allows talk/activity unrelated to subject
	4. Conducts beginning/ending review				▓▓▓		4.
	5. Questions: academic comprehension / lesson development	a. Single Factual (Domain 5.0)					5a. Allows unison response
		b. Requires Analysis/ reasons					5b. Poses multiple questions asked as one
							5c. Poses nonacademic questions/ nonacademic procedural questions
	6. Recognizes response/amplifies/gives correct feedback						6. Ignores student or responses/expresses sarcasm, disgust, harshness
	7. Gives specific academic praise						7. Uses general, nonspecific praise
	8. Provides for practice						8. Extends discourse, changes topic with no practice
	9. Gives directions/assigns/checks comprehension of homework, seatwork, assignments /gives feedback						9. Gives inadequate directions on homework/no feedback
	10. Circulates and assists students						10. Remains at desk/circulates inadequately
4.0 PRESENTATION OF SUBJECT MATTER	11. Treats concepts – definition/attributes/examples/ nonexamples						11. Gives definition or examples only
	12. Discusses cause-effect/uses linking words/applies law or principle						12. Discusses either cause or effect only/ uses no linking words
	13. States and applies academic rule						13. Does not state or does not apply academic rule
	14. Develops criteria and evidence to value judgment						14. States value judgment with no criteria or evidence
5.0 COMMUNICATIONS VERBAL AND NONVERBAL	15. Emphasizes important points				▓▓▓		15.
	16. Expresses enthusiasm verbally/challenges students				▓▓▓		16.
	17.		▓▓▓		▓▓▓		17. Uses vague/scrambled discourse
	18.		▓▓▓		▓▓▓		18. Uses loud, grating, high-pitched, monotone, or inaudible talk
	19. Uses body behavior that shows interest – smiles, gestures						19. Frowns, deadpan or lethargic
2.0 MANAGEMENT OF STUDENT CONDUCT	20. Stops misconduct						20. Delays desist/doesn't stop misconduct/desist punitively
	21. Maintains instructional momentum						21. Loses momentum – Fragments nonacademic directions, over dwells

OBSERVER'S NOTES: _____

_____

An Equal Opportunity Employer

FC-710-1010   (Rev. 02/12/08)

APPENDIX C
DATA COLLECTION SHEET

	District Implimentation Variables								Teacher Evaluation Ratings			
	1	2	3	4	5	6	7					
County	Delibrate Practice	Number of Formals	Scope of Formal Elements	Pre & Post Conferences	Informal Observations	Classroom Walkthroughs	Domains 2, 3, & 4		(1) Highly Effective	(2) Effecitve	(3) Need Imp & Dev	(4) Unsat.
Bradford	y	n	y	y	y	y	y		15	144	40	1
Broward	n	n	y	n	y	n	n		1524	13766	148	1
Calhoun	n	n	y	y	y	n	y		9	149	6	1
Charlotte	n	n	n	y	n	n	y		216	705	23	3
Collier	n	n	y	y	y	y	y		0	3391	0	0
Franklin									4	83	1	0
Gadsden	n	y	n	y	y	n	y		90	257	20	0
Gilchrist	y	n	y	n	n	y	y		94	57	0	1
Indian River	n	n	y	y	y	y	y		0	0	0	0
Jackson	n	n	y	y	y	y	y		20	442	17	0
Lafayette	n	n	n	y	y	y	y		47	22	1	0
Lake	n	n	y	n	y	n	n		100	2746	82	2
Leon	y	n	y	y	y	y	y		1930	234	4	0
Martin	n	n	y	y	y	n	y		748	390	0	0
Nassau	n	n	y	n	y	y	y		252	421	5	0
Orange	n	n	n	y	n	n	y		776	10195	122	1
Osceola	n	n	n	n	n	n	y		1119	2456	34	3
Palm Beach	n	n	y	n	n	y	n		4331	7090	21	4
Pasco	y	n	y	y	y	y	y		184	3651	67	0
Putnam	n	y	y	y	y	y	y		142	477	5	0
Santa Rosa									999	653	8	1
Seminole									2871	1337	32	0
St. Johns	y	n	y	y	y	n	y		800	1018	8	0
St. Lucie	n	n	y	y	y	n	n		3	63	33	0
Union	y	n	n	y	y	y	y		63	95	6	0

LIST OF REFERENCES

Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives.* Cincinnati, OH: South-Western.

Cubberley, E. P. (1929). *Public school administration.* Boston: Houghton Mifflin. Buckley, R., & Wiese, D. S. (1998). The evolution of the performance appraisal process.

Journal of Management History, 4(3), 233-249.

Ericson, K., Krampe, R., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychology Review, 100*(3), 363-406.

Farkas, S., Johnson, J., & Duffett, A. (2003). *Stand by me: What teachers really think about unions, merit pay, and other professional matters.* New York: Public Agenda.

Florida Department of Education. (2014a). Educator recruitment, development, & retention. Retrieved from http://www.fldoe.org/profdev/pdf/teipm.pdf

Florida Department of Education. (2014b). District performance evaluation systems. Retrieved from http://www.fldoe.org/profdev/pdf/EduEvalRatingsMarch14.pdf Florida Department of Education. (2014c). Educator recruitment, development, & retention. Retrieved from http://www.fldoe.org/profdev/pdf/teipm.pdf Florida Department of Education. (2014d). Approved instructional performance evaluation systems directory. Retrieved from http://www.fldoe.org/profdev/instructional.asp

Flowers, A. (2013). *A study of the Marzano teacher evaluation model and student*

achievement at 24 elementary schools in a large urban school district in Central

Florida. (Unpublished Doctoral Dissertation). University of Central Florida, Orlando, FL.

Glatthorn, A. (1984). *Differentiated supervision.* Alexandria, VA: ASCD.

Goldhammer, R. (1969) *Clinical supervision: Special methods for the supervision of teachers.* New York, NY: Holt, Rinehart & Winston.

Grubb, T. (2007). Performance appraisal reappraised: It's not all positive. *Journal of Human Resources Education, 1*(1), 1-22.

Haefele, D. L. (2004). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education, 7*, 21-31.

Hattie, J. A. C. (1992). Measuring the effects of schooling. *Australian Journal of Education, 36*(1), 5-13.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement.* Abingdon, England: Routledge.

Hunter, M. (1980). Six types of supervisory conferences. *Educational Leadership, 37*(5), 408-412.

Kersten, T., & Israel, M. (2005). Teacher evaluation: Principals insights and suggestions for improvement. *Planning and Changing, 35*, 47-67.

Kirkpatrick, D. L. (2006). Training and performance appraisal--are they related? *Training and Development, 60*(9), 44-45.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on

performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychology Bulletin, 119*(2), 254.

Kosmoski, G. (2000). *Supervision* (2nd ed.). Mequon, WI: Stylex.

Lawler, E. E., Benson, G. S., & McDemmot, M. (2012). What makes performance appraisals effective? *Compensation and Benefits Review, 44*(4), 191-200.

Livingston D., & Livingston P. (2012). *Teacher development toolkit for the Marzano teacher evaluation model.* Englewood, CO: Marzano Research Laboratory. Retrieved from

http://www.marzanoresearch.com/documents/TeacherDevelopmentToolkit.pdf

Manasa, K. L., & Reddy, N. (2009). Role of training in improving performance. *IUP Journal of Soft Skills, 3*(3/4), 72-80.

Manoharan, T. R., & Muralidharan, C. (2012). A composite model for employees' performance appraisal and improvement. *European Journal of Training and Development, 36*(4), 448-480.

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

Mayo, E. (1946). *The human problems of an industrial civilization.* Cambridge, MA: Harvard University Press.

Micceri, T., Peterson, D, & Smith, B. O. (1985). Measurement of teacher performance: A study of instrument development. *Teaching and Teacher Education. 1*(1), 63-67.

National Council on Teacher Quality (2010). *2009 State teacher policy yearbook:*

National summary 2009. Retrieved from: http://www.nctq.org/reports.do

National Council on Teacher Quality (2012). *State of the states 2012: Teacher*

effectiveness policies. Retrieved from: http://www.nctq.org/reports.do

Pichler, S. (2012). The social context of performance appraisal and appraisal reactions: A

meta-analysis. *Human Resource Management, 51*(5), 709-732.

Reavis, C.A. (1976). Clinical supervision: A timely approach. *Educational Leadership,*

33, 360-363.

Rothman, R., & Toch, T. (2008). *Rush to judgment: Teacher evaluation in public*

education. Education Sector Reports, American Institutes for Research.

Selvarajan, T. T., & Cloninger, P. A. (2012). Can performance appraisals motivate

employees to improve performance? *International Journal of Human Resource*

Management, 23(15), 3063-3084

Sergiovanni, T. J., & Starratt, R. J. (2007) *Supervision: A Redefinition*. Boston, MA:

McGraw-Hill

Student Success Act, SB. 736, (2011).

Taylor, F. W. (1911). *The principles of scientific management and testimony before the*

special house committee. New York, NY: Harper & Row.

U.S. Department of Education. (2009). *Race to the top grant.* Retrieved from

http://www2.ed.gov/programs/racetothetop/index.html

U.S. Department of Education. (2012). *Race to the top: Year one final results.* Retrieved

from http://www2.ed.gov/programs/racetothetop/phase1-applications/score-

summary.pdf

Wiese, D. S., & Buckley, M. R. (1998). The Evolution of the performance appraisal process. *Journal of management history, 4*(3), 233-249.