

University of Central Florida

STARS

Electronic Theses and Dissertations

2014

An analysis of the difference between assessed instructional practices, value-added measures, and learning gains of secondary reading teachers

Ronald Fritz

University of Central Florida



Part of the [Educational Leadership Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Fritz, Ronald, "An analysis of the difference between assessed instructional practices, value-added measures, and learning gains of secondary reading teachers" (2014). *Electronic Theses and Dissertations*. 4636.

<https://stars.library.ucf.edu/etd/4636>

AN ANALYSIS OF THE DIFFERENCE AMONG ASSESSED INSTRUCTIONAL
PRACTICES, VALUE-ADDED MEASURES, AND LEARNING GAINS OF
SECONDARY READING TEACHERS

by

RONALD SCOTT FRITZ

B. A. University of South Florida, 1991

M. A. University of South Florida, 1994

ED. S. Nova Southeastern University, 1996

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

Fall Term
2014

Major Professor: Rosemarye Taylor

© 2014 Ronald Scott Fritz

ABSTRACT

The purpose of this study was to analyze the relationship that exists among instructional practice scores, value-added measures, and learning gains for the Large Urban School District (LUSD) among reading teachers in Grades 6-12 for the 2012-2013 school year. In addition, this study was also conducted to address the perceptions of secondary school principals and assistant principals regarding the relationship of the three variables and their relevance in making personnel decisions.

Quantitative data were obtained from school district databases for observation scores, value-added measures, and standardized assessment achievement data to determine the relationship among the variables. In addition, a perception survey was completed by secondary school principals and assistant principals. With a sample size of 138, the survey yielded a return rate of 84%. The data were analyzed to determine actual relationships among instructional practice scores, value-added measures, and learning gains, as well as how the participants perceived each variable in isolation.

The literature review supported the findings regarding the inflation of observation scores by school administrators. The only significant relationship (.48) existed between value-added measures and learning gains. Likewise, school leaders in the LUSD believed that their observations were important in analyzing the needs of their teachers and should be included in summative evaluations. Learning gains were supported by the participants more than instructional practice scores and value-added measures as evidence of effective instruction. The literature review revealed past program evaluation

studies regarding the Florida Comprehensive Achievement Test (FCAT) as a valid measure for assessing grade level benchmarks.

The information in this study is valuable and suggests that continued professional learning for school leaders regarding classroom observations to improve inter-rater reliability is needed. Likewise, school leaders would benefit from understanding the relationships that exist among instructional practice scores, value-added measures, and learning gains to drive conversations with teachers regarding rigorous instruction.

To Nana

ACKNOWLEDGMENTS

The experience and knowledge that I have received from completing this dissertation has been the highpoint of my professional career as an educator. The required commitment could not have been maintained without the unending support of my soul mate, Anna Ruth. Likewise, I would like to acknowledge my children, Natasha and Reese, for being my constant reminder of what is really important in life. Furthermore, I must recognize my mom, my dad, and my brothers, David and Jason, for teaching me the meaning of responsibility and dedication.

I will never be able to express enough gratitude to Dr. Rosemarye Taylor for her continuous support throughout my studies and research. I would also like to thank the members of my committee, Dr. Lee Baldwin, Dr. Walter Doherty, and Dr. Mary Pat Kennedy. Your guidance and support have made me a better researcher and educational leader.

I must also thank my professional mentors, Dr. Michael A. Grego and Dr. Barbara M. Jenkins. Both of these superintendents unselfishly encouraged me to complete this degree for my own professional growth.

The process of completing this dissertation has reaffirmed my belief that human beings are capable of more than they believe possible given the right support. My action item is to serve as that support for others with their dreams.

TABLE OF CONTENTS

LIST OF TABLES	x
CHAPTER 1 THE PROBLEM AND ITS CLARIFYING COMPONENTS.....	1
Introduction.....	1
Problem Statement	4
Purpose Statement.....	5
Definition of Terms.....	6
Categories	6
Progression of Teacher Performance	8
Summative Evaluation Ratings	9
Operational Definitions.....	10
Assumptions.....	13
Limitations	13
Delimitations.....	14
Conceptual Framework	14
Research Questions and Hypothesis	20
Methodology	21
Population and Sample	21
Data Collection	22
Data Analysis	22
Significance of the Study	24
Organization of the Study	26
CHAPTER 2 REVIEW OF LITERATURE	27
Introduction.....	27
Observations, Value-added Models, and Learning Gains	29
Principal Observations of Teachers	30
Value-added Models	36
Florida Comprehensive Assessment Test	48
School Leaders' Perception	52
Principals' Decision-Making Using the Data	61
The Principal as the Leader for Improving Teacher Quality	61
Teacher Renewal.....	68
Summary	74
CHAPTER 3 METHODOLOGY	79
Introduction.....	79
Purpose of the Study	81
Research Questions	81
Participants.....	82
Instrumentation	84

Data Collection	86
Data Collection Procedures.....	86
Quantitative Data Collection.....	88
Qualitative Data Collection.....	91
Data Analysis	91
Quantitative Data Analysis	93
Qualitative Data Analysis	93
Summary	94
CHAPTER 4 RESULTS	96
Introduction.....	96
Research Question 1	96
Research Question 2	105
Research Question 3	114
Instructional Practice Scores.....	114
Open-ended Responses: Instructional Practice Scores	117
Value-added Scores	119
Open-ended Responses for Value-Added Scores	121
Learning Gains.....	123
Open-ended Responses for Learning Gains.....	126
Administrators' Open-Ended Responses: Strategies to Improve the Effectiveness of Reading Teachers	127
Summary	130
CHAPTER 5 SUMMARY, DISCUSSION, AND IMPLICATIONS	132
Introduction.....	132
Summary of the Study	132
Discussion of the Findings.....	135
Research Question 1	135
Research Question 2	141
Research Question 3	146
Qualitative Themes	149
Implications and Recommendations for Practice	151
Recommendations for Future Research	154
Summary	159
APPENDIX A SCHOOL DISTRICT PERMISSION TO CONDUCT THE STUDY	161
APPENDIX B UNIVERSITY OF CENTRAL FLORIDA INSTITUTIONAL REVIEW BOARD APPROVAL.....	163
APPENDIX C COMMUNICATIONS WITH PARTICIPANTS	165
APPENDIX D INSTRUCTIONAL PRACTICE, VALUE-ADDED MEASURE, SECONDARY READING LEARNING GAINS SURVEY	170

REFERENCES	181
------------------	-----

LIST OF TABLES

Table 1	Research Questions, Sources of Data, and Variables	24
Table 2	Observation Rubric: Categories of Leaders and Descriptors.....	35
Table 3	Florida Comprehensive Assessment Test (FCAT) Achievement Levels	49
Table 4	Frequencies and Percentages: 2011-12 LUSD Teachers Assigned to Performance Categories (N=12, 558)	60
Table 5	Racial Composition of Participants: Principals and Assistant Principals for Instruction (API)	83
Table 6	Research Questions, Sources of Data, Analysis, and Variables	92
Table 7	Descriptive Statistics for All Reading Teachers (N=883)	98
Table 8	Descriptive Statistics for Middle School Reading Teachers (N=522).....	100
Table 9	Descriptive Statistics for High School Reading Teachers (N=361)	102
Table 10	Pearson r Correlational Analysis Among Teacher Effectiveness Measures: All Teachers (N=883)	103
Table 11	Correlational Analysis Among Teacher Effectiveness Measures: Middle School Teachers (N=522)	104
Table 12	Correlational Analysis Among Teacher Effectiveness Measures: High School Teachers (N=361)	104
Table 13	Student Demographics: Participating Principals' Schools (N=116)	106
Table 14	Personal and Professional Demographics: Participating Principals (N=116)	107
Table 15	Administrators' Opinions: Usefulness of Instructional Practice Scores (N=106).....	109
Table 16	Administrators' Opinions: Usefulness of Value-added Scores (N=98)	112
Table 17	Administrators' Opinions: Usefulness of Learning Gains Scores (N=97) ...	113
Table 18	Administrators' Opinions: Use of Instructional Practice Scores to Make Decisions (N=106)	116

Table 19	Administrators' Confidence in the Use of Instructional Practice Scores (N=106).....	117
Table 20	Administrators' Open-ended Responses: Instructional Practice Scores (N=13)	118
Table 21	Administrators' Opinions: Use of Value-added Scores to Make Decisions (N=98).....	121
Table 22	Administrators' Open-ended Responses: Value-added Scores (N=20).....	122
Table 23	Administrators' Opinions: Use of Learning Gains to Make Decisions (N=97)	125
Table 24	Administrators' Open-ended Responses: Learning Gains Scores (N=27) ...	127
Table 25	Administrators' Open-ended Responses: Strategies to Improve Reading Teachers' Effectiveness (N=52)	128

CHAPTER 1 THE PROBLEM AND ITS CLARIFYING COMPONENTS

Introduction

The pendulum of education has oscillated many times over the last century. An increased concern at the national, state, and local levels has narrowed the focus of what constitutes effective teaching, and heavy emphasis has been placed on student growth models (Ravitch, 2010). Under the American Recovery and Reinvestment Act (ARRA) of 2009, the federal government funded an educational initiative for \$4.35 billion known as Race to the Top. This program was designed as a competitive four-year grant to provide innovative reform in the area of education (U.S. Department of Education, 2012). A key focus of the ARRA grant was to accomplish the following: (a) close the achievement gaps of various subgroups of students, (b) improve graduation rates, and (c) increase the number of students prepared for college and careers (U.S. Department of Education, 2012). Specifically, the Race to the Top initiative focused reform within four areas:

(a) the adoption of rigorous standards and assessments to prepare students for college and career, (b) the building of data systems to communicate student success to principals and teachers, (c) the recruit, retain, and award of highly effective instructors and leaders, and (d) the transformation of the lowest-performing schools. (U.S. Department of Education, 2012, p. 2)

As a recipient of the grant, the state of Florida received \$700 million to accomplish these goals. In addition, Florida's reform agenda included the passing of the

Student Success Act for the purpose of aligning the state strategic plan with the Race to the Top initiative. The Student Success Act solidified the reform path in Florida. As such, it (a) established a thorough evaluation system for teachers and principals based on student growth and teacher observations; (b) linked compensation for teachers and principals to the final results on performance evaluations; and (c) eliminated tenure for teachers with the exception of those educators who had been awarded tenure prior to its passage. In addition, the Student Success Act required that all students be assessed for learning growth in every course by the year 2014-2015 (Florida State Statute 1012.34, 2012). This study was to be conducted to analyze the relationship between the evaluation of teachers by their principals and the value-added model (VAM) score based on student learning gains for the 2012-2013 school year.

In the first year of Race to the Top, all Florida school districts adopted a teacher evaluation system and began the transition towards the Common Core State Standards (CCSS). Preparation for administrators and teachers on the new models was initiated, and school districts began practicing with the new observation instruments. However, efforts to accomplish the first-year goals were delayed due to leadership changes and lack of qualified staff at the state level (U.S. Department of Education, 2012).

In the second year, the CCSS were initiated in kindergarten. Principals began using the formative observation tools to provide feedback to teachers regarding the rating of their instructional practices. Initially, school districts were mandated to use the value-added measure assigned to teachers for a minimum 40% of the summative evaluation (Florida State Statute 1012.34, 2012). In the state of Florida, most school districts opted

to place emphasis on the instructional practices portion of the summative evaluation (U.S. Department of Education, 2013).

In the third year, the state continued with the implementation of the CCSS in kindergarten and first grade. In addition, local educational agency consortiums began assisting the state with developing hard to measure assessment items for courses such as physical education, music, and foreign language that would ultimately be used to gauge student learning, as well as impact teacher evaluations (U.S. Department of Education, 2013). Consequently, for the large urban school district in which this study was conducted, the substance of this legislation required the production of 1,422 assessments ready to demonstrate student growth by the fall of 2014 (Florida Organization of Instructional Leaders, 2012).

To prepare school leaders for the anticipated changes, a series of inter-rater reliability professional learnings were conducted for the purpose of improving principals' skills in rating teachers' instructional practices using the school district-adopted observation instrument. To a large extent, scores have shown that administrators evaluated a majority of teachers as Effective or Highly Effective in the area of instructional practices (Florida Department of Education, 2013c).

By 2013-2014 under Race to the Top, all school districts in the state of Florida were expected to implement the CCSS in Grades K-2 for mathematics and English language arts. In 2014-2015, CCSS was scheduled to be implemented in kindergarten through Grade 12 with the emergence of a national assessment being administered in Grades 3-11 (U.S. Department of Education, 2013).

Problem Statement

There were 138 site-based principals and assistant principals in Large Urban School District (LUSD) in Florida which was the target of this study. All participants were administrators in either middle or high school. Site administrators were responsible for evaluating the instructional staff under their purview. For the 2012-2013 school year, 60% of teachers' evaluations were comprised of the scores earned on the instructional practices portion of the evaluation instrument, scores obtained through observations using iObservation for effective teaching strategies (Large Urban School District, 2011).

As of 2014 principals and assistant principals were required to conduct formal and informal observations of teachers to provide feedback regarding their levels of proficiency with the one of 10 specified design questions and related elements. School administrators were able to rate teachers based on evidence observed regarding the implementation of the learning goal through a system of teacher evaluation scales (Marzano, 2007) with corresponding numerical values as follows: (a) Not Using = 0, (b) Beginning = 1, (c) Developing = 2, (d) Applying = 3, or (e) Innovating = 4. After each observation, the teacher's rating on the element was banked and accumulated for a summative analysis. Annually, by May 1, all instructional personnel received one of the following preliminary ratings: (a) Highly Effective (b) Effective, (c) Needs Improvement (d) Developing, and, (e) Unsatisfactory (Large Urban School District, 2011). In late July, the state of Florida annually assigned each teacher a value-added score in accordance with Florida State Statute 1012.34, (2012) based on a growth model for those students who were in attendance at selected points in the school year as determined by Full Time

Equivalent surveys 2 and 3 each year. The value-added score was used for the remaining 40% of the teacher's overall evaluation. Beginning in the 2014-2015 school year, the value-added score was to be universally established and applied across the state of Florida, equating to 50% of teachers' summative evaluations (Florida State Statute 1012.34, 2012).

The problem addressed in the present study was related to the lack of prior research on the relationship between the three elements of a teacher's final evaluation: (a) the value-added measure, (b) the instructional practices score, and (c) the actual student learning gains obtained as measured by the FCAT Reading . These three elements may or may not be aligned, and they may or may not be valid measures of a teacher's performance. Furthermore, those evaluating teachers may not have alignment in their assessment of classroom practices.

Purpose Statement

The purpose of the analysis was to determine the relationship that exists among the instructional practices portion of the teachers' summative evaluations conducted by principals and assistant principals, the assigned valued-added score based on student growth in the area of reading for Grades 6 through 8 and 9 through 12, and the learning gains earned by these teachers.

Definition of Terms

For this analysis, it was critical to define terminology for the reader to comprehend the significance of the school administrator's portion of the evaluation process as related to the assigned results of the value-added measure and learning gains earned for related instructional personnel in LUSD. Thus, terminology is presented by: (a) broad categories as established by the LUSD; (b) the progression of teacher performance on the evaluation instrument; (c) summative evaluation ratings; and (d) operational definitions for terminology pertinent to this study.

Categories

Categories are broad classifications that have been established for teachers based on individual experience and expertise. These categories were created through collective bargaining with the localized teachers union and approved by the department of education (Florida State Statute 1012.34, 2012).

Category 1. This category consists of teachers with 0-2 years of teaching experience. Teachers with zero years are defined as first year teachers and those with two years are in their third year of the profession (Large Urban School District, 2013a).

Category 2A. This category consists of teachers at least in their fourth year (Large Urban School District, 2013a).

Category 2B. This classification is for experienced teachers who have at least three years of teaching experience with one of the following applications: (a) new hire to the district, (b) teachers who have been assigned to instruct a new subject that is different

from their previous assignment, (c) teachers employed at a school with a different student population from previous year, or (d) teachers who earned between a 2.0 and a 2.4 for an instructional practice score for the previous year. Teachers meeting any one of these criteria may request that the principal move them to this category. Likewise, principals may assign teachers to this grouping based on their meeting one of the above criteria (Large Urban School District, 2013a).

Category 3. This group is for teachers who have been rated ineffective in the classroom either through observable behaviors with instructional practice or have been assigned valued added measures that do not exhibit the required learning growth. Category 3 teachers receive augmented support and feedback as well as being placed on a Professional Improvement Plan (PIP) that includes additional observations. Once the PIP is completed, the principal reassigns the teacher to the original classification. If Category 3 teachers are unsuccessful in fulfilling the requirements of the PIP, they may receive an overall Needs Improvement or Unsatisfactory rating on the final evaluation (Large Urban School District, 2013a).

Category 4. This classification is for teachers who were originally in another category; however, there is a lack of sufficient data in all four domains for instructional practice to be evaluated equitably due to one of the following conditions: (a) significant leave of absence or (b) beginning employment after February 15 of the school year (Large Urban School District, 2013a).

Progression of Teacher Performance

Terminology associated with the progression of teacher performance on the observation instrument is defined in sequential order as follows:

Not Using. According to Marzano (2007), when a principal or an assistant principal rates a teacher as Not Using during a formative observation, there is no evidence of a learning goal or instructional strategy. This is the lowest point on the scale for the formative observations of teachers performed by principals and assistant principals to gauge instructional practice (Marzano, 2007).

Beginning. According to Marzano (2007), when a teacher receives a Beginning rating on a formative observation for instructional practice, it is due to utilizing the strategy incorrectly or because of omitted necessities, e.g., cite a couple. This scale indicates that there is a great extent of growth required (Marzano, 2007).

Developing for Instructional Practice. According to Marzano (2007), when teachers receive a Developing rating on the formative instrument, they have presented a clear learning goal and a scale. However, the monitoring of student understanding of the learning has not been addressed (Marzano, 2007).

Applying. According to Marzano (2007), when teachers receive an Applying rating on a formative observation, it can be assumed that the teacher had a specific learning goal and accompanying scale that described the levels of performance as well as monitored students' understanding of the learning goal. In addition, teachers who receive an Applying rating from their principals or assistant principals should have shown significant documented evidence (Marzano, 2007).

Innovating. According to Marzano (2007), teachers who receive an Innovating rating on their formative observations are at the highest category a teacher can receive. When a principal or assistant principal observes teachers and rates them as Innovating, it can be assumed that the following have been documented: (a) specific learning goals are provided with accompanying scales; (b) the teacher is monitoring for understanding of the learning by the students; and (c) instruction is adapted and modified with new strategies to meet the needs of individual students and situations (Marzano, 2007).

Summative Evaluation Ratings

Five summative evaluation ratings can be assigned. They represent an amalgam of the previously defined formative ratings:

Unsatisfactory. For teachers who earn a summative evaluation of 60% for instructional practice and 40% for the valued-added measure, the overall score will fall between 1.00 and 1.49. This equates to Not Using on the scale utilized for instructional practice (Large Urban School District, 2013a).

Needs Improvement. Teachers who receive an overall evaluation of Needs Improvement must be assigned to categories 2A, or 2B. Their cumulative range of performance is between 1.50 and 2.49. This equates to the Developing rating within the Marzano model (Large Urban School District, 2013a).

Developing. The difference between Needs Improvement and Developing is the category in which the teacher is placed. This classification is for Category 1 teachers

only and would correlate to the corresponding scale for instructional practice (Large Urban School District, 2013a).

Effective. For teachers to receive an Effective summative evaluation, scores range from 2.50-3.49. The parallel scale is Applying. Effective teachers have students with at least one grade level of academic growth in an academic year (Large Urban School District, 2013a).

Highly Effective. This grouping is for teachers earning the highest classification who have a summative evaluation of 3.50-4.00 with an applicable scale of Innovating. Highly effective teachers have students who earn one and one half grade levels of academic growth over the course of the school year (Large Urban School District, 2013a).

Operational Definitions

Administrators Observations. Periodic and regularly scheduled visits by school administrators in order to witness instruction first hand for the purpose of determining areas of strengths and weaknesses with individual teacher's delivery of instruction (Danielson, 2011).

Design Questions. According to Marzano (2007), design questions "represent a logical planning sequence for effective instructional design" (p. 7). Each question is used to guide the instruction and to build on sound pedagogical practice.

Florida Comprehensive Assessment Test (FCAT). The state of Florida's annual assessment which measures student success with the Next Generation Sunshine State

Standards, includes assessments in reading for Grades 3 through 10, mathematics for Grades 3 through 8, writing for Grades 4, 8, and 10, and Science for Grades 5 and 8 (Florida Department of Education, 2013d)

Focused feedback. After observations of instructional practice, principals and assistant principals engage teachers in discussion based on the evidence they collected during the observation. This discussion targets the design question, the learning goal, and the effectiveness of delivery as determined by the level of the scale given to the observation (Marzano, 2007).

Instructional Practices. This term is defined by the observations that principals and assistant principals conduct of teachers regarding their levels of competence in delivering instruction. These observations can be both formal and informal and are based on one of the 10 design questions in the Marzano model (Marzano, 2007).

Learning Gains. Learning gains are defined as a year's worth of knowledge learned over the same period in time. This is determined in the state of Florida by comparing the scores of students' prior performance on the Florida Comprehensive Assessment Test (FCAT) to their outcome of the current year's assessment. In the state of Florida, there are three different methods to determine learning gains:

- (a) Improve one or more FCAT 2.0 achievement levels (e.g., from 1-2, 2-3, 3-4, or 4-5) or Florida Alternate Assessment (FAA) performance levels (for students with significant cognitive disabilities);
- (b) Maintain a proficient achievement level on the FCAT 2.0 or FAA (at least level 3 for the FCAT 2.0, level 4 for the FAA) without decreasing a level; or

(c) Demonstrate more than one year's growth when remaining in achievement level 1 or 2 on the FCAT 2.0 (or when remaining at performance level 1, 2, or 3 for the FAA) for both years. Under this alternative, one year's growth on the FCAT 2.0 is defined in terms of the difference between a student's current year and prior year FCAT 2.0 vertical scale score. To make learning gains, students who remain at level 2 on the FCAT 2.0 have to score at least one point beyond a year's expected growth. Students who remain at level 1 have to score at least two points beyond a year's expected growth. FAA students who remained at performance level 1, 2, or 3 are credited with gains if their score improves by at least five (5) points (raw points) compared with the prior year's score. (Florida Department of Education, 2013d, p. 11)

Professional Improvement Plan (PIP). This is a process in which struggling teachers receive assistance in order to increase their instructional practice in the classroom. A strategic plan is created in order to focus on strategies in one of the four domains measured by the Marzano model (Large Urban School District, 2013a)

Value-added measure. Value-added measure is the metric assigned to specific teachers based on the growth in the learning of the students they taught during a specified period of time (Ravitch, 2010). The difference between the predicted performance and the actual performance represents the value-added by the teacher's instruction (Florida Department of Education, 2014b).

Assumptions

1. In this analysis, it was assumed that all principals and assistant principals had received required professional learning offered by the school district for utilizing the Marzano teacher effectiveness instrument.
2. It was realistic to suppose that all principals and assistant principals understood the following operational definitions: (a) Not Using, (b) Beginning, (c) Developing, (d) Applying, and (e) Innovating.
3. It was assumed that all principals and assistant principals comprehended the summative evaluation classifications as follows: (a) Unsatisfactory, (b) Needs Improvement, (c) Developing, (d) Effective, and (e) Highly Effective.

Limitations

This study had the following limitations:

1. The instructional practice scores being correlated to the valued-added measures were only in the area of reading.
2. The analysis included only secondary schools in which the principals and assistant principals were responsible for evaluating teachers in the area of instructional practice.
3. The value-added measures assigned to teachers were based only on a standardized assessment administered in the state of Florida.
4. The rater reliability of the principal and assistant principal, in using the observation instrument, had not been assessed at the time of this analysis.

5. In the Large Urban School District that was the focus of this study, other members of the administrative staff, in addition to the building principal, often conduct instructional practice observations on teachers.

Delimitations

For this study, the researcher analyzed data from only one school district with one evaluation system. This limited the generalizability of the instructional practice results across other populations of principals in other school districts. Furthermore, because the state standardized test was to be converted to an assessment instrument to measure the newly adopted Common Core State Standards, the analysis only informed the school district how closely correlated the instructional practice scores were with the value-added measures regarding the obsolete assessment tool.

Conceptual Framework

The roles of the principal and assistant principal have clearly changed in the first decade of the 21st century. Previously, school leaders' responsibilities were related more to management than to leadership. Good school-based administrators were not necessarily instructional leaders. Federal initiatives, however, have emphasized the importance of effective principals in augmenting the teaching and learning in schools. State policies across the nation have been dictated from published reports such as Alger's work with Nebraska's Platte Institute, as cited in Marzano, Toth, and Schooling (2012) and recommendations indicating the need for "better evaluation measurement systems for

teachers and principals, higher accountability, and an absolute focus on improved educator effectiveness and student learning” (p. 5). According to a study conducted by the Wallace Foundation (2013), school leaders who address educational challenges in isolation, often fail to foster a learning environment that is necessary in order to augment student learning. To maximize learning opportunities in a school, the principal or assistant principal must be able to create conditions in which effective teaching is the priority and universally supported by all other areas of the school (Wallace Foundation, 2013). This change in focus has resulted from (a) the need to create successful workers who are prepared to compete in an international economy and (b) the heightened awareness of the discernible achievement gap among various groups of students. Intense efforts at all levels of education have been launched to increase standards and develop rigorous learning goals to meet the needs of all students (Wallace Foundation, 2013).

Serving as instructional leaders has presented some different challenges for school leaders, depending on the school levels they lead. According to Shelton (2011) elementary principals have had better success with becoming content area experts than those at the secondary level. This is largely due to the volume of curriculum addressed in middle and high schools. Therefore, at the secondary school level, effective principals must utilize their leadership skills to construct administrative teams of experts to assist with furthering their academic vision. Consequently, these teams spend purposeful time in classrooms observing and providing feedback to teachers (Shelton, 2011)

According to Cogan, as cited in Marzano and Simms (2013), traditionally, supervisors observed teachers in the classroom, identified areas needing improvement,

shared the information with the teachers, and communicated how they wanted the changes to occur. In this historical model, teachers and school site administrators progressed through the obligatory stages of the classroom observation without engaging in collegial inquiry (Marzano & Simms, 2013).

As a result of high stakes accountability in the field of education, present-day principals and assistant principals increasingly have been expected to serve as the instructional leader for their teachers. Consequently, it has become necessary for their roles to shift to perform an advisory or coaching function. Gauthier and Giber, as cited in Marzano and Simms (2013), discovered the origin of the term coaching in the field of transportation. They defined coaching as assisting individuals to move from where they currently are to where they need to go. Joyce and Showers (2002), stated that coaching by school leaders was the most effective method of assisting teachers with the transfer of pedagogy and content to the classroom. Furthermore, Kretlow and Bartholomew (2010) found “strong evidence for the effectiveness of coaching in promoting the fidelity of evidence-based practices” (pp. 292-293).

Several conditions must occur to assist teachers in improving instructional strategies. First, trust must exist between the person providing the feedback and the recipient. Schools, in which teachers feel a great sense of trust, have positive relationships with their administration. Furthermore, teachers who have faith in their site administration, openly acknowledge them as effective leaders that are supportive of professional learning and prioritize the learning of the students within their school (Bryk & Schneider, 2002).

For feedback to produce the desired results, Brockband and McGill, as cited in Marzano and Simms (2013), established the following criteria for leaders: (a) be clear, (b) own the feedback, (c) start with the positive, (d) be specific, (e) focus on the behavior, (f) be descriptive. Feedback should also include strategies that teachers are using correctly, and teachers should be active participants in the process by having some choice in the necessary adjustments. The ultimate goal is to allow teachers to become innovative and more cognitive of effective practices and their own instruction. Consequently, effective feedback enables educators to establish obtainable goals and track their own performance (Hattie, 2009).

It would be difficult to find an educator who would argue against the concept that quality teaching is the single most important variable in student learning. To be an effective instructional leader, principals must be able to recognize high quality lessons when they observe them based on research from meta-analysis. Likewise, to perform at high levels, teachers must possess a high degree of competency with Webb's Depth of Knowledge and cognitive rigor (Hess, Carlock, Jones, & Walkup, 2009). In their review of a study conducted by The Standards Company, Hess et al., reported that the majority of teaching at the third-grade level for English language arts and mathematics occurred within the three lowest levels of the revised Bloom's Taxonomy and the first two levels of Webb's Depth of Knowledge. Conversely, when students were assessed, the majority of the assessment items were on higher levels than was the instruction students received. In the state of Florida, 50% to 70% of the items on the Grade 3 FCAT Reading were considered to be moderately complex. An additional 5% to 15% of the questions were

considered highly complex test items (Florida Department of Education, 2013a).

Therefore, to improve the quality of teaching, principals and assistant principals must understand and be able to communicate the components of cognitive science as well as have a thorough understanding of the required shifts in instruction to prepare students for the inevitable standardized assessments (Koedinger, Corbett, & Perfetti, 2012).

Another challenge facing teachers in Florida has been the requirement of an embedded value-added measure as part of their summative evaluations. The original concept of this growth model was developed by Dr. William Sanders at the University of Tennessee who was interested in the extent to which teachers contributed to the learning gains of their students (Ravitch, 2010). Sanders' model was solely statistical in nature and did not involve any actual observations of teachers. According to Sanders and Horn (1994), systems can be developed to neutralize factors affecting student outcomes including such issues as: (a) mobility of students, (b) modes of teaching, (c) altering teaching assignments, and (d) regression to the mean. The underlying conclusion of Sanders and Horn was that effective teachers tend to be so with students from all achievement levels. As the idea grew, emphasis was placed on computation, rather than on methodology of instruction (Ravitch, 2010).

Low value-added scores based on poor student performance have been linked with ineffective teachers. In fact, students with ineffective teachers, as delineated by value-added measures, are more likely to become pregnant, not attend college, and have lower earning professions later in life (Winters, 2012). Besides pure pedagogical purposes, value-added measures can be utilized as a data point for conducting human

resource functions. It has been a common practice in the American education system to award tenure status to teachers for surviving a minimal number of years in the profession. Typically, this period is three years and has traditionally not been based on performance. There is little doubt that children have been inadequately prepared by a system that secures the employment of ineffective educators (Winters, 2012). In recent years, school districts have adopted value-added measures to remove tenure status from those teachers who receive below satisfactory performance ratings (Briggs, 2011). Proponents of the value-added models argue that quantifying teacher performance exposes ineffective teachers more easily (Glazerman, Loeb, Goldhaber, Staiger, Raudenbush, & Whitehurst, 2010). Likewise, results from a Florida study reveal that the value-added measures of non-tenured teachers early in their careers are statistically significant predictors of future performance (Winter, 2012). While those opposing growth models claim the opposite, there is no doubt that student growth models afford leaders more confidence in executing decisions regarding staff retention and promotion (Glazerman et al., 2010).

In the state of Florida, the value-added measure has been applied to Florida Comprehensive Assessment Test (FCAT) learning gains for each teacher and the students they instructed over a fixed period of time (American Institute for Research, 2010). There are multiple methods to calculate the student learning gains on FCAT as explained in the operational definitions presented earlier in this chapter (Florida Department of Education, 2013d). When considering the value-added measure in the context of Florida's FCAT, several factors must be considered. First, the amount of growth measured varies significantly across different achievement ranges within specific grade

levels. Second, the average growth varies significantly across grades. Finally, the average scores on the assessment fluctuate within specific grade levels across years (American Institute for Research, 2010).

Research Questions and Hypothesis

1. To what extent was there a relationship among each administrator's evaluation of teachers' instructional practices, the value-added measure, and the learning gains assigned to teachers from learning gains as measured by Florida Comprehensive Achievement Test (FCAT) Reading for the following grades: (a) 6 through 8 and (b) 9 through 12 in an urban school district for the school year 2012-2013?

H₀. There is no relationship between instructional practice scores attained through administrators' observation, the value-added measures, and the learning gains assigned to teachers based on the learning gains of their students in reading as determined by the FCAT in Grades 6 through 8 or 9 through 12.

H_A. There is a relationship among the instructional practice scores attained through administrators' observations, the value-added measures assigned to teachers, and the learning gains of their students in reading as determined by the FCAT in Grades 6 through 8 or 9 through 12.

2. What factors do middle, and high school principals and assistant principals believe contribute to the relationships among the instructional practice ratings, the value-added measures, and the learning gains?
3. To what extent do principals report using the instructional practices score, learning gains, or VAM scores to make personnel or instructional decisions?

Methodology

Population and Sample

The population for the present study consisted of 275 middle and high school principals and assistant principals in Large Urban School District in the state of Florida for the 2012-2013 school year. The purposeful sample consisted of 138 principals and assistant principals for instruction who completed instructional practice scores for reading teachers in Grades 6 through 8 and 9 through 12.

It was necessary to identify eligible secondary reading teachers so that data could be obtained regarding instructional practice scores, value-added measures, and learning gains. The population of secondary reading teachers in which data was secured consisted of 955 educators from the school district. The sample comprised of teachers with scores for all three variables. Consequently, 883 eligible reading teachers from the school district were included in the analysis.

Data Collection

Instructional practices scores were collected through the iObservation instrument utilized by the Large Urban School District. The researcher had access to the school district data and received written permission to use it (Appendix A). Approval to conduct the study was also received from the Institutional Review Board of the University of Central Florida prior to the initiating the research study (Appendix B). Value-added measures have been delivered to school districts from the state of Florida annually and have been maintained in the iObservation . Finally, learning gains for each reading teacher were available in the Educational Data Warehouse maintained by the school district.

Data Analysis

Research Question 1 was used to investigate the extent to which there was a difference between each administrator's evaluation of teachers' instructional practices and the value-added measures assigned to teachers from learning gains as measured by FCAT Reading for the following grade levels: (a) 6 through 8 and (b) 9 through 12 in an urban school district. Quantitative data from summative results of instructional practice scores for all teachers in corresponding grades in reading were correlated with the value-added measures assigned based on learning grains with FCAT Reading using the Pearson product-moment coefficient of correlation (Pearson r). The instructional practice scores served as the independent variable. The value-added measures, learning gains for the

secondary school reading teachers, were the dependent variables. Scores for both measures were compared through scatterplots.

To respond to Research Question 2, as to those factors that middle and high school principals and assistant principals believed contribute to the differences among the instructional practice ratings and value-added scores, survey data were analyzed. Using a five point Likert-type scale, school administrators from each school level completed one fixed response survey querying them about (a) instructional practice scores, (b) valued-added measures, and (c) learning gains. Measures of central tendency to include mean and standard deviation were calculated.

Responses for Research Question 3 were open ended and designed to elicit responses from principals and assistant principals regarding the extent that instructional practice scores, valued-added measures, and learning gains were used to make personnel decisions for the 2013-2014 school year. Responses were categorized by type of data used to improve reading as (a) learning gains, (b) value-added measures, and (c) instructional practice scores. In addition, participants were asked to share personal beliefs regarding improving the effectiveness of reading teachers.

Table 1 presents the research questions, the sources of data that were used to respond to each question, and the variables that were employed in the analysis of the data.

Table 1

Research Questions, Sources of Data, and Variables

Research Questions	Source of Data	Variables
To what extent was there a relationship between each administrator's evaluation of teachers' instructional practices, learning gains, and the value-added measures assigned to teachers as measured by FCAT Reading for the following grade levels: (a) 6-8 and (b) 9-12 in an urban school district?	Instructional practice scores Assigned value-added measures Learning Gains	Independent: instructional practice scores Dependent: Value-added measures and learning gains
What factors do middle and high school principals and assistant principals believe contribute to the relationships among the instructional practice rating learning gains, and value-added measures?	Instructional practice, learning gains, and value-added measure survey of principals and assistant principals	Independent: Instructional practice scores Dependent: Value-added measures and learning gains
To what extent do principals report using the instructional practices score, learning gains, or VAM scores to make personnel or instructional decisions?	Open Ended question as part of the survey for principals and assistant principals	Independent: Instructional practice scores Dependent: Value-added measures and learning gains

Significance of the Study

An analysis of the original intent of federal legislation and the actual application of the initiatives is essential in order to gauge the progress achieved in improving instruction in Large Urban School District. As evaluation systems evolve to encompass research based instructional practices, thereby satisfying requirements of state legislation, it is expected that teacher performance will improve, and so will student achievement

(Marzano et al., 2012). If this holds true, the ability of principals and assistant principals to gauge effective instructional practice of teachers during formative observations and sharing meaningful feedback is paramount to improving student achievement (Marzano & Simms, 2013)

Other studies have received great notoriety. As an example, the Measures of Effective Teaching (MET) project released its initial analysis that the teacher evaluation system was essentially ineffective, with 98% of teachers receiving a “Satisfactory” or higher rating. This project, funded by the Bill and Melinda Gates Foundation, sought to connect teacher observations with student performance. It included 3,000 teachers from the following school districts: (a) Charlotte-Mecklenburg, (b) Dallas, (c) Denver, (d) Hillsborough County, (e) New York City, and (f) Memphis. All participants were English language arts teachers in Grades 4 and 8, and the observation instrument was designed by Danielson (Bill & Melinda Gates Foundation, January 2012).

With the increased attention given to principals and assistant principals serving as instructional leaders and the evidence supporting research-based strategies in the classroom, the present study was expected to reveal the extent to which intent correlated to actual application. Revealing the relationship of the three measures (instructional practice, value-added, and learning gains) was intended to provide decision makers with research for further policy development. This research was intended to add to the body of knowledge on the relationship of the three measures: instructional practice, value-added measures, and learning gains. It should provide decision-makers with a foundation for further policy development.

Organization of the Study

This study is described and reported in five chapters. Chapter 1 has provided an overview of the study. Chapter 2 contains a review of the literature and research related to relevant aspects of the problem. Chapters 3 and 4 are comprised of a description of the methods and procedures used to conduct the study and the analysis of the data, respectively. The fifth and final chapter presents a summary of the data, implications for policy and practice, and recommendations for future research.

CHAPTER 2 REVIEW OF LITERATURE

Introduction

With the initiation of the Race to the Top grant and the alignment of Florida's Student Success Act, the evaluation of school based administrative and instructional staff has been linked to the learning gains of students within the respective schools (Florida State Statute, 1012.34, 2012). For administrators, the scope of their evaluation has begun to include the learning growth for all students in all content areas within their schools. For instructional staff, the value of their instruction for their particular students is measured by the percentages of learning gains and students' earning proficiency on the Florida Comprehensive Assessment Test (FCAT) (Florida State Statute 1008.22, 2012). In addition, principals are responsible for conducting observations of instructional delivery throughout the school year in order to generate a summary of the individual teacher's daily performance. These two measures calculated together determine the summative value for final annual evaluation (Florida State Statute 1012.34, 2012).

The database search for this review of the literature was completed using resources from the University of Central Florida. The data bases included: Education Full Text, ERIC, Dissertations & Theses Full Text, Professional Development Collection, and PsychInfo. The key words used to search the databases consisted of: principal observations, teacher observations, teacher effectiveness, academic achievement, instructional practice, classroom observations, teacher feedback, professional learning, principal coaching, mentoring, value-added, value-added assessment, value-added measures, value-added models, learning gains, secondary reading gains, academic gains, academic

achievement, Florida Comprehensive Assessment Test, principal's perceptions, principal's attitudes, principal's beliefs, administrator beliefs, principal's decision-making, personnel decisions, hiring effective teachers, retention of effective teachers, tenure, and dismissal of ineffective teachers. The researcher reviewed literature online as well as printed journals including: *Journal of Research on Leadership Education*, *Educational Administration Quarterly*, *Educational Leadership*, *Educational Forum*, *Journal of Educational Research*, *Journal of Educational and Behavioral Statistics*, *Journal of Education for Students Placed at Risk (JESPAR)*, *Teacher and Teacher Education*, *Journal of Educational Administration*, *Middle Grades Research Journal*, *Journal of Staff Development*, *Educational Review*, *Journal for School Leadership*, *Journal of Human Resources*, *Journal of Labor Economics*, *American Economic Review*, *Educational Policy Analysis Archives*, *Personnel Psychology*, and *Journal for Personnel Evaluation in Education*.

The Internet was also used to locate websites for the literature review. Websites that were accessed were those of the U.S. Department of Education, the Florida Department of Education, Florida Center for Reading Research, Florida State Statutes, American Institute of Research, the Wallace Foundation, MET, the College of Education at the University of Washington, and the Center for Educational Research.

The review of literature consists of three sections, each addressing the literature pertaining to a specific research question for the study. In section 1, relevant literature pertaining to (a) classroom observations by administrators, (b) the various implementations of value-added models used for performance assessments and their perceived use, and (c) information regarding the Florida Comprehensive Assessment Test (FCAT) for middle and high school in the area of secondary reading was reviewed. In section 2, literature focused on

the perceptions of school leaders regarding the factors of importance regarding classroom observations, the value-added by classroom teachers, and the achievement scores earned by students on standardized test scores. In section 3, literature regarding how principals utilize the data in order to make personnel decisions for their schools was reviewed.

Observations, Value-added Models, and Learning Gains

In 1917, William Connor, State Superintendent for Michigan, wrote, “If a teacher is rated at all, she should be rated, not by the clothes she wears, or the methods she chooses, but by the results she secures” (Connor, 1917, p. 338). By the 1950s, unions were in full swing and began to address the teacher evaluation process. Most states operated under a collective bargaining decree limiting the power of principals and school administrators. In 1984, Texas implemented the Texas Teacher Appraisal System (TTAS). This system allowed administrators to review the activities of a teacher. In the 1990s, the state began using standardized assessments to collect data on how teachers were affecting student learning. In Boston Public Schools, the collective bargaining process drove the changes in the teacher evaluation system. The union claimed that school administrators inaccurately recorded their performance and pushed for reform. The National Commission on Teaching and America’s Future, supported by the Rockefeller and Carnegie Foundations concluded that the reform of elementary and secondary schools must begin with revamping the teaching profession (Darling-Hammond, 1996). In 1998, Cincinnati Public Schools worked collaboratively with the teachers’ union to establish an evaluation system based on incentive pay, professional

learning, and certification. The committee created a peer educator evaluation model. The revised model proposed that the evaluation for teachers should be comprised of five observations of instruction. Two of these observations were to be completed by the principal or the assistant principal and three by a peer teacher evaluator (Cincinnati Public Schools Collective Bargaining Agreement, 2000). Further evolutions of teacher evaluation systems defined effective teaching by the scores earned on standardized assessments and the classroom observations performed by school principals. Positive results were rewarded and sanctions were issued for poor performance (Darling-Hammond (2009).

Principal Observations of Teachers

Effective teachers have some common traits: (a) they possess a deep understanding of content knowledge, (b) they connect what is to be learned with student's prior knowledge, (c) they create effective scaffolding and learning supports, (d) they use strategies that assist students with drawing connections and apply their new knowledge, (e) they assess student learning and adjust their teaching to meet the students' needs, (f) they provide clear constant feedback, and (g) they develop effective classroom management in which students feel membership to the group (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012)

According to the findings in the Measures of Effective Teaching (MET) Project, principals should evaluate teachers through regular classroom observations, as well as by the learning gains of the students. The researchers suggested a minimum of four

observations from the principal with a research-based rubric (Measures of Effective Teaching, 2010). Other researchers have suggested that it is daily instruction that produces augmentation in student learning. According to Marshall (2012), principals should conduct at least 10 unannounced observations of 10-15 minutes duration for each classroom teacher. In addition, she expressed the belief that all observations should be conducted by the same administrator in order to ensure consistency. Other researchers have suggested that principals should use various rubrics to compensate for the inadequacies of individual tools in order to provide more accurate feedback for teacher evaluations (Kane & Cantrell, 2012). In addition, Marshall advocated that principals should make efforts to observe the same teacher at different times of the lessons and on various days of the week. Afterwards, the principal and the teacher should have a face-to-face meeting for the purpose of providing additional coaching on the observed instruction (Marshall, 2012).

Good teaching is supported by time spent planning for instruction. Observations allow principals to witness the interactive work of the teacher with their students. Regardless of how good the planning is, if the delivery in the classroom is deficient, the overall instruction cannot be considered effective (Marzano, 2007). Determining what administrators are targeting during the observation is often reliant on the instructional framework adopted by the state or school district. Observers must acquire sufficient training in order to conduct meaningful observations on teachers. In fact, many states require that administrators receive certification to complete observations (Marshall, 2012). When observing, it is important for principals to write what they actual see and

hear. This can be extremely difficult for educators, as opinion and interpretation are easy ways for observers to contaminate actual evidence. Once observers collect the evidence, their role is to compare the collected information against an established rubric that gauges levels of performance. This results in observers refocusing their interest from what happened in the lesson to what the events, activities, and actions mean (Danielson, 2011).

Holding professional conversations with teachers after observations is a critical function of improving instruction. Administrators must be skilled at interacting with teachers in order to encourage them to begin reflecting on their instruction (Marzano, 2007). Because teaching is complicated, all instruction can be improved. This requires that the teacher be an active participant in the process as a learner. In addition, observers must be receptive to altering their interpretation of the evidence if teachers present a convincing argument for an alternative perspective. Because all states require some type of observation of teachers, systems must be designed to observe clear standards of practice in a natural state (Danielson, 2011).

A review of the literature did not yield a definitive answer in regard to the accuracy of principals' observations at gauging high quality instruction. In 2009, the New Teacher Project's Widget Effect Study revealed the following information regarding principal evaluations of teachers in Chicago: (a) 25,332 teachers were rated as Superior, (b) 9,176 were Excellent, (c) 2,232 were Satisfactory, and (d) 149 were Unsatisfactory. These results indicate that almost all teachers were rated as highly effective or effective. Researchers determined a significant reason for the inflated evaluations of the teachers was due to the announced formal observations by the school principal. In addition, the

findings in this study revealed that a majority of school principals were unable to identify high quality instruction accurately. Furthermore, if the principal did identify areas of concern, the abilities of the leaders to respond in a meaningful way were inconsistent. The impact of the inept observation skills of the principals encouraged poor teachers to continue ineffective practices while simultaneously failing to highlight the more effective instructors. As a result, the principals tended to render the same evaluations for all teachers.

The administrative practice of observing teachers has been widely known and practiced in education (Weisberg, Sexton, Mulhern, & Keeling, 2009). Peterson (2000) concluded that the average practice of teacher observation was not an accurate representation of the actual instruction in the classroom. Likewise, Danielson and McGreal (2000) criticized teacher evaluation systems as having little relevance compared to actual student learning. According to Kane and Staiger (2012), a valid observation instrument must be aligned with the outcomes of the students. Subsequently, Weisberg et al. further suggested that unless principals' observations reflect the day to day instruction accurately, they are virtually useless to serving as an instrument to improve teacher effectiveness.

Conversely, there are school districts that have had success in aligning principals' observations of the teachers with overall student performance. In a pilot study in Chicago, the Danielson observation tool was employed by trained principals and observers. School leaders, as well as peer observers, were given extensive professional learning on the instructional framework. Both groups became proficient in the use of the

observation instrument and practiced using inter-rater reliability training sessions. Overall, the observation instrument was deemed a valid measure of teacher effectiveness when compared to the value-added measures for the same group of students. For example, teachers who received low observation scores generally had low value-added scores at the end of the year. Likewise, those educators who had high scores on the observation tool generally had correlated value-added scores. However, reliability was an issue for principals and observers. Although both observers and principals tended to have similar ratings for low performing teachers, principals scored proficient teachers as distinguished more often than did observers. This lack of consistency by principals led teachers to believe that the framework was subjective (Sartain, Stoelinga, & Brown, 2011).

In a study conducted by Fink (2012), 2,207 principals from 42 school districts were assessed regarding their skills as classroom observers. The participants were rated on their ability to observe, analyze, and respond in meaningful ways as the instructional leaders for their schools. The rubric was based on a 4-point scale across five dimensions. The results of the study expressed as aggregated data for the five dimensions were as follows: (a) purpose, 1.48; (b) student engagement, 2.04; (c) curriculum and pedagogy, 1.93; (d) assessment for student learning, 1.73; and (e) classroom environment and culture, 1.70. The categories to which the participants were assigned based on the 4-point rubric and the descriptors for each of the categories are displayed in Table 2.

Table 2

Observation Rubric: Categories of Leaders and Descriptors

Category of Leader	Descriptors
Novice Instructional Leader	Does not think about key concepts when observing classrooms.
Emerging Instructional Leader	Identifies elements related to key concepts. Uses relevant terminology. May ask questions without elaboration and offers suggestions without justification.
Developing Instructional Leader	Elaborates responses with examples. Demonstrates basic understanding. Offers alternatives to teaching decisions.
Expert Instructional Leader	Critically analyzes observed lessons. Conveys clear vision. Communicates evidence and examples. Demonstrates pedagogical expertise. Links questions to evidence of student learning.

Source. Fink (2012)

Findings indicate that elementary principals scored higher than middle or high school principals. Furthermore, length of experience was not a predictor of high scores. In fact, the most significant predictor among the principals was expertise with instruction pedagogy. Finally, differentiation did exist between suburban, rural, and urban principals with the first group scoring slightly higher. However, Fink (2012) was unable to conclude whether this difference was significant.

Value-added Models

Emerging interests in aligning student learning gains with educational accountability has inspired unparalleled attempts to embed high stakes assessments in the evaluation of individual teachers and schools, and state and district policies have conceptualized student learning as scores obtained through standardized assessments. Though researchers have produced some data supporting the accuracy and the stability of the value specific teachers add to the impact on student learning, little research has been conducted through a comprehensive and systemic process over time and across course offerings to calculate teacher effect (Newton, Darling-Hammond, Haertel, & Thomas, 2010).

Conceptually, the value-added measure promises to quantify the added impact of the teacher in terms of their students' learning gains (Sanders & Horn, 1998). VAM is attractive because it offers a way to disentangle the effects of teachers from those of other uncontrolled factors such as: (a) student demographics, (b) socio-economic characteristics, (c) family education, (d) language background, and (e) neighborhood environment. Despite the appeal, VAMs have been designed to show student growth when students have been randomly assigned to teachers. This is seldom the actual case. Students have often been purposively placed with specific teachers. It has been extremely common for a given teacher to be assigned a disproportionate number of students with greater challenges ((Newton et al., 2010). Significant instability exists in teachers' value-added scores. They have varied from class to class and year to year and

have been rooted in changes in student characteristics that have then been associated with changes in the teacher ratings (Jerald, 2009).

Because VAMs have been used to make high stake decisions, states and school districts must determine the competing values to include within their frameworks. As models are considered, policymakers must analyze and separate school influences from teacher effects with significant consideration being given to VAMs that have a history of stability of the teacher impact across time. The final consideration involves determining how the curriculum will be structured so that the adopted VAM is applicable to daily instruction. This can be particularly tedious at the secondary level due to the specificity of individual course content (Newton et al., 2010).

At the time of the present study, 40 states and the District of Columbia were using, piloting, or in the process of developing a growth model in order to measure student learning and teacher effectiveness. The most popular models include: (a) the Educational Value-Added Assessment System (EVAAS), (b) the Student Growth Percentiles (SGP) model, and (c) the Value-Added Research Center (VARC) model. For states to adopt a value-added model, data systems must be able to connect individual students with their teachers of record. According to the Data Quality Campaign (DQC) in 2011, only seven states (14%) were able to connect students to the teacher of record (Collins & Amrein-Beardsley, 2012).

The remainder of this review of value-added models will focus on the implementation of various models in leading states around the nation. Included are

models which have been implemented in Tennessee, Louisiana, Texas, California, and Florida.

Tennessee

The Tennessee Value-Added Assessment System (TVAAS) was utilized for the purpose of providing student achievement scores that were void of biases typically associated with standardized tests. The TVASS was comprised of a statistical mixed model method in order to perform multivariate analysis of student performance. Consequently, the data can be aggregated to a particular classroom in a school (Sanders & Horn, 1994).

Because the model omits statistical controls for social economic status (SES), demographic, and other determinants of academic achievement, researchers have been very critical of the model. Controlling for SES and demographics is complex due to the relationship of these variables and teacher quality. Results have indicated that restricting for SES and demographic factors at the student level is unbiased when analyzing teacher effects with the TVAAS. Because student achievement data is not always annually available, stringent requirements serve to isolate the measures to be added to the value of the teacher effectiveness (Ballou, Sanders, & Wright, 2004).

Studies have been conducted using the data from TVAAS on the CTBS-5 assessment in order to determine if transformation efforts were effective. In 2001, researchers reviewed fourth and fifth grade TVAAS in both reading and mathematics in schools that had participated in reform efforts and those that abstained. Across the board,

all reformed schools improved significantly as measured by the TVASS (Ross, Wang & Alberg, 2001).

Another study conducted in Tennessee identified 40 national board certified educators in Grades 3 through 8 and analyzed their students' learning gains as measured by the Tennessee Value-Added Assessment System (TVAAS). Of the 40 identified teachers, 16 failed to be classified as effective due to their students earning inadequate learning gains in one or more content areas including mathematics, reading, or language arts or they failed to meet the benchmark for three consecutive years (Stone, 2002).

Louisiana

In reviewing the value-added models in the state of Louisiana, it is important to consider teacher preparation programs. Variance among the effectiveness of teacher preparation programs across postsecondary institutions has been determined to be significant, and this has impacted the quality of the teachers assigned to students in local schools. Like many states and school districts, inadequate data systems have been the greatest challenge in analyzing the effectiveness of the teacher preparation programs. Another hurdle regarding the VAM employed for assessing the effects of new teacher programs is that the scores measured have not necessarily been isolated to the year of instruction for the student-teachers. Despite the obstacles, supporters have suggested that the value-added model provides data that can be used to improve current teacher evaluation practices. Because of perceived success, in 2000, Louisiana's Blue Ribbon Commission for Teacher Quality determined criteria to recruit, select, prepare, and

support new teachers. Consequently, the Louisiana Board of Regents required all universities to redesign their teacher preparation programs based on the 60 recommendations. Subsequently, a teacher preparation accountability system was created. The three main components of the system included: (a) academic performance of the student teachers including passage of the PRAXIS; (b) quantity of program completers; and (c) achievement scores of student taught by the new teachers. The value-added analysis was employed to control for variability among teacher preparation programs. It also afforded researchers the opportunity to measure teacher effectiveness in specific content areas (Gansle, Noell, & Burns, 2012).

Texas

In 2010, the superintendent and school board of the Houston Independent School District invested in a software product called Education Value-Added Assessment System (EVAAS). This district initiative was aligned with the federal goal of having the United States lead the world in college completion by 2020. As a result of the new teacher evaluation system, a case study was conducted to review the factors that resulted in the termination of four teachers (identified as Teachers A, B, C, and D) within the school district. All four teachers were inner city elementary teachers with an average of 11.8 years of teaching experience and 7.5 years in Houston. Their peers had nominated teacher A and teacher C as Teachers of the Year in previous years. Both Teachers A and C also received merit pay the year prior to their dismissals. The scenarios of the dismissals of the four teachers are instructive.

Teacher A always received high marks on her evaluation from her principal. Her value-added scores displayed equal positive and negative impacts on student learning. Teacher B revealed a negative two years of student growth and one positive year. Her value-added scores mirrored the observations of her principal. Her last year of teaching was deemed her most effective year, as indicated by the value-added score assigned to her and her corresponding principal observations. Consequently, teacher B displayed significant improvement in both realms. Teacher C's value-added scores exhibited that her involvement with her students negatively impacted their learning for each of the three years considered. Teacher C taught the highest needs students in the school including those significantly overage for their grade level. Teacher D also showed both positive and negative impacts on student learning. During the year of the evaluation that resulted in her termination, teacher D received an influx of English language learners into her class. Her value-added scores were among the bottom for the school district. Teacher A, B, and D resigned. Teacher C challenged the termination and pursued her case in court. The verdict from the magistrate ruled in favor of teacher C on the grounds that the data provided by the standardized assessment was inconsistent and not statistically significant enough to result in termination (Holloway-Libell, Armrein-Beardsley, & Collins, 2012).

The Houston study produced evidence that under their VAM, teachers with large numbers of English language learners and students with exceptionalities were found to have lower value-added scores. Likewise, teachers with gifted students displayed little student growth because their students were annually earning top scores on the standardized assessments. In addition to these results, unintended consequences have

caused teachers in Houston to shy away from teaching assignments involving the neediest students. Similarly, teachers have actively pursued employment in grade levels where the value-added scores are easier to show student growth (Holloway-Libell et al., 2012).

California

Secondary value-added models have been criticized largely because of the variation with content among courses (Sawchuk, 2012). A study was conducted of secondary English language arts and mathematics teachers from six high schools with 3,500 students based on the variation in student test scores on the California Standards Tests (CSTs). The strengths of the study included: (a) matching of students with their teachers at the course level which afforded the researchers the ability to rank teachers across courses, (b) matching of teachers with students who were present the entire academic year, and (c) studying educators at the high school level which highlighted several barriers for utilizing value-added models. The limitations included: (a) insufficient data systems were available to measure student growth across the state, (b) data were missing in high mobility areas which correlated to schools in low socio-economic communities, and (c) measurement of student growth through end of course examinations for content areas, e.g., Biology and Chemistry, did not have clear learning progressions such as Biology and Chemistry.

The summative data regarding the VAM model used revealed that the most visible difference in the teacher effect was related to whether or not student demographics and school influence were controlled. Furthermore, the use of multi-year

models did not alter the effects significantly. The summary of these findings suggested that teachers who had great ratios of advantaged students had higher rankings as compared to some of their peers. Furthermore, for teachers who taught both the upper and lower track of students, their VAM scores were significantly higher for the upper group of students. Correlations from the study suggested that even in the most complicated models, there was a significant amount of variation regarding the value a teacher contributed and the student characteristics (Newton et al., 2010).

In another research study involving a large urban school district in the state of California, value-added data were obtained from four cohorts of fifth graders. For each cohort, achievement scores were utilized for Grades 3, 4, and 5 for English language arts and mathematics. The study involved 3,651 teachers and 161,811 students from 469 schools. During this study, researchers analyzed the effects of three value-added models. Controls were included in the models. In total, value-added estimates were evaluated across 14 conditions (Kersting, Chen, & Stigler, 2012).

There were several interesting findings in the Kersting et al. (2012) study. First, approximately one-third of all teachers in the study had students who increased their performance across all 14 conditions being measured. Second, researchers concluded that the student sample size was most impactful (32%) on individual teacher reclassification. Next, approximately two-thirds of all teachers remained in the same classification for all 14 conditions. Finally, the top value-added scores were more consistent than the bottom scores. The researchers concluded that the majority of the teachers maintained their classification because standards of errors that adjusted for those

teachers in either the low or high group were used in the data analyses (Kersting et al., 2012).

Florida

As was stated previously, Florida was obligated to utilize a value-added model as a requirement of the Student Success Act as well as its participation in Race to the Top grant funding (Florida Department of Education, 2014b). Further, the Student Success Act specifically stated that instructional evaluations would be based on the performance of each teacher's assigned students, instructional practice, and professional job responsibilities (American Institute for Research, 2011).

In the state of Florida, eight committees were created to carry out the work associated with Race to the Top. Over 200 people applied to serve on the committee that would select Florida's value-added model (Florida Department of Education, 2014d). Ultimately, 27 individuals (teachers, school administrators, district level administrators, postsecondary teachers, parents, and community members) reviewed various growth models (Florida Department of Education, 2014b). This committee became known as the Student Growth Implementation Committee (SGIC) (Florida Department of Education, 2014). Members represented the diverse culture and various regions of the state. The chair of the committee was Sam Foerster, Associate Superintendent in Putnam County (Florida Department of Education, 2014d). The role of the committee was to solicit feedback from stakeholders and provide a recommendation to the Commissioner of

Education regarding the proposed growth model (Florida Department of Education, 2014b).

To begin the work, the American Institute for Research (AIR) selected eight current value-added models being employed around the nation for the committee to consider: (a) the Sander's model, (b) the Rand model, (c) the hybrid model 1, (d) the hybrid model 2, (e) the Meyer model, (f) the hybrid model 3, (g) the differences model, and (h) the Colorado model. AIR conglomerated the eight value-added models into two main classes as typical path models and covariate adjustment models. The typical path models proposed that teachers and students can alter their learning over time. An important feature of these models was that they did not precisely regulate for prior achievement. Conversely, for the covariate adjustment models, prior student test scores were directly controlled as predictors of student performance. Through much deliberation and analysis, the SGIC selected an approach from a class of covariate adjustment models. The model recommended by the SGIC assumed that when students were receiving instruction from a teacher with average effectiveness, they would earn achievement scores that paralleled their peer students with similar performances and like characteristics. It was anticipated that a positive effect with the chosen model would produce an increase in student performance from the predicted values (Florida Department of Education, 2014b). The Commissioner of Education approved the recommendation of the SGIC and contracted with AIR to develop and support the model (Florida Department of Education, 2014d).

The results from various content areas such as reading and mathematics can be demonstrated to a small extent, and a distinct regression for reading and mathematics or jointly can be used (Florida Department of Education, 2014b). Though the latter produces some statistical challenges, the estimated teacher effects from a joint and marginal model were correlated at greater than .99; and provisional variances of the teacher impact displayed only insignificant differences (Lockwood, McCaffrey, Mariano, & Setodji (2007). An issue that can have a significant impact on covariate models results pertains to the impact of error among the predictor variables. Various covariate models use instrumental variables to control for error. In cases of high stakes decision making, ignoring the potential for this error can augment skepticism among critics (Florida Department of Education, 2014b).

The covariate adjustment model adopted by the state of Florida includes two years of previous scores on the FCAT with the exception of fourth grade due to only having one year's data available in order to calculate predictive scores. The model adopted by the state of Florida was designed to be neutralizer for the following: (a) student characteristics, (b) classroom characteristics, and (c) school characteristics (Florida Department of Education, 2014c).

In order to understand the various components of Florida's value-added model, it is necessary to describe the predictor variables that the model takes into account. The Florida model specifically analyzes each student's FCAT scores for reading and mathematics in Grades 3-10. The dependent variable is the most current reading or mathematics score on the FCAT. The predictor variables included in the Florida model

are the same for both reading and mathematics. First, the model determines the number of subject-relevant courses that are linked to an FCAT in which individual students are enrolled according to the Florida course code directory. Next, two prior years of FCAT achievement scores for reading and mathematics are secured. Then, the Florida model determines the disability status for each student as well as English language learner status. Some students may have multiple variables. Other factors included in the Florida value-added model include: (a) gifted status, (b) mobility, (c) attendance, (d) difference from modal age, (e) class size, and (f) homogeneity of prior test scores. Consequently, “The Florida VAM applied to the FCAT data decomposes total variation in achievement into three orthogonal components: variation between schools, variation between teachers within the school, and variance between students within the classroom” (Florida Department of Education, 2014b, p. 4).

The variation between schools measures the amount of student learning that is expected for all students in each school that differs from the statewide expectation. Regarding the variation between teachers within the school, this element analyzes the growth of the students amongst the teachers within the same school. Finally, the last factor reviews the individual student growth for each of the students assigned to specific teachers.

There are several advantages of the Florida covariate adjustment model. Advocates for the adopted model have touted the following:

- (a) Teachers teach classes of students who enter with different levels of proficiency and possibly different student characteristics;

- (b) Value-added models ‘level the playing field’ by accounting for differences in the proficiency and characteristics of students assigned to teachers;
- (c) Value-added models are designed to mitigate the influences of differences among the entering classes so that schools and teachers do not have advantages or disadvantages simply as a result of the students who attend a school or are assigned to a class. (Florida Department of Education, 2014d, p. 8)

Florida Comprehensive Assessment Test

According to Florida State Statute 1008.22 (2012), “The primary purpose of the student assessment program is to provide student achievement and learning gains to students, parents, teachers, school administrators, and school district staff. This data is used by districts to improve instruction” (p. 1). The Florida Comprehensive Assessment Test (FCAT) assesses English language arts, mathematic proficiency and learning gains. For the purpose of this literature review, reading has been the content area of interest.

Beginning in 1996, the FCAT began to be used to measure student performance on established benchmarks. Table 3 shows the resulting five achievement levels that were established

Table 3

Florida Comprehensive Assessment Test (FCAT) Achievement Levels

Achievement Levels	Descriptors
Level 5	The student has had success with the most challenging content and has answered a majority of the questions correctly including the most complex.
Level 4	The student has had success with the most challenging content and has answered a majority of the questions correctly.
Level 3	The student had partial success with the challenge questions and is considered proficient.
Level 2	The student has limited success with challenging content.
Level 1	The student has little success with challenging content.

Source. Florida Department of Education, 2013e.

Scale scores for each level are determined to show where each student falls within the achievement and grade level. Learning gains are defined as a year's worth of knowledge learned over the same period in time. This has been determined in the state of Florida by comparing the scores of students' prior performance on the Florida Comprehensive Assessment Test (FCAT) to their results of the current year's assessment. In the state of Florida, there have been three different methods to determine learning gains: students may (a) increase one or more achievement levels as dictated on the FCAT, or the same for students with exceptionalities on the Florida Alternative Assessment (FAA), (b) maintain a proficient achievement level on the FCAT 2.0 or FAA without decreasing a level, or (c) improve more than one year's growth when remaining

in achievement level 1 or 2 on the FCAT or when remaining at performance level 1, 2, or 3 for the FAA in for both years (Florida Department of Education, 2013d).

Cognitive factors change significantly for students from Grades 3-10. In the area of reading, the greatest change is the significance behind the activity. For example, the third grade FCAT measures fluency but the 10th grade FCAT measures verbal reasoning. The average level 1 student in third grade can read only 54 words per minute. In contrast, the average level 1 student in 10th grade reads 130 words per minute. Though the third-grade level 1 student is focused on fluency, the 10th-grade level 1 student has made great gains in fluency and has become more focused on the meaning of the words. The FCAT mirrors this progression. On the third-grade FCAT, approximately 30% of the questions require complex thinking as opposed to 70% on the 10th-grade FCAT. From third to 10th grade, higher order thinking skills augment exponentially (Torgesen, Nettles, Howard, & Winterbottom, 2004).

Schatschneider, Buck, Torgesen, and Wagner (2004) conducted a study to isolate the reading, cognitive, and linguistic skills that impact the performance of students on the FCAT in Grades 3, 7, and 10. Primarily, they wanted to determine the major deficiencies for level 1 and 2 students in order to assist in the planning of intervention strategies. The sample consisted of 200 students from the 2003 FCAT. Participant demographics included: (a) 41% white, (b) 39% African American, (c) 17% Hispanic, and (d) 2% Asian. In addition 56% were female and 44% were male. The schools that were chosen to participate were from Tallahassee, Tampa, and Ft. Lauderdale. The tests given to the samples measured the following five broad reading areas: (a) verbal knowledge and

reasoning, (b) text reading fluency, (c) phonemic decoding efficiency, (d) non-verbal reasoning, and (e) working memory.

The instrument used with the sample for the third- and seventh-grade students was the SAT 9 or FCAT normed referenced tests. The analysis for the level 1 and 2 third-grade students displayed that these students were significantly below their proficient peers in fluency. Furthermore, the level 1 students were significantly deficient in phonemic decoding. However, verbal reasoning was at the 42% level. The seventh-grade level 1 and 2 students also experienced difficulties with fluency, and their phonemic decoding was also weak compared to their proficient peers. In addition, the seventh-grade level 1 students were further behind in verbal knowledge and reasoning than the level 1 third-grade students. Worth noting, level 2 students in seventh grade scored at the 51st percentile on the normed-referenced test. With respect to the level 1 and 2 10th-grade students, fluency and phonemic decoding continued to be an area of declining performance as compared to proficient peers. The gap in verbal knowledge and reasoning widened for level 1 students but narrowed for level 2 students since taking the seventh grade FCAT. The researchers concluded that Florida's struggling students were losing substantial ground between Grades 3 and 10 regarding their verbal knowledge and reasoning skills (Schatschneider et al., 2004).

Validity studies regarding the FCAT have been conducted since its inception to verify the content of the assessment. In 2010, the Buros Center for Testing conducted an operational check of the 2010-2011 FCAT for 10th-grade reading. The results of the study were compared to the internal results obtained by the Florida Department of

Education. The analysis included item calibration, scaling, and the equating of the 10th-grade FCAT Reading tests. The results from the external party determined that the entire process of the FCAT was well organized and the Florida Department of Education used reasonable and justifiable calibration, scaling, and equating conclusions (Chin, Shaw, Dwyer, McCormick, & Geisinger, 2010).

School Leaders' Perception

It has become very apparent that America's principals have been increasingly burdened with the responsibility of ensuring that their teachers are providing high quality instruction to their students. As previously noted, 21st century principals and assistant principals have been required to observe teachers in their daily instruction and evaluate them based on the acquired evidence. In addition, in many states school leaders utilize the value-added scores for teachers as well as data from student outcomes on standardized assessments in order to complete summative appraisals. Policy makers and school district leaders must analyze the perceptions of their school leaders regarding what constitutes teacher effectiveness as compared to the actual data supporting teacher performance (Gordan, Meadows, & Dyal, 2001). Some scholars would argue that the future of public education depends on present day school leaders' ability to accurately judge the quality of instruction (Medley & Coker, 1987). The section of the review focuses on the discernments of school leaders regarding the factors that promote teacher effectiveness.

Friedman, Friedman, and Markow's 2008 longitudinal study can be used to illustrate how the role of student achievement has augmented in importance in recent years. The purpose of this study was to conduct a nationwide perception inventory of school principals and their assistants to determine their overall satisfaction with their roles as school leaders. The intent of the research was to isolate predictors that corresponded with the job satisfaction of the typical American administrator. Through a series of focus groups, the researchers began formulating targeted questions as early as 1993. A total of 431 school leaders completed the 136-item survey, resulting in a 69% return rate. The survey produced 12 indexes of satisfaction. The top correlations for satisfaction with the role of school principal were: (a) central office support (.89); (b) three indicators, i.e., parental, school board, and superintendent support (.88); and (c) decision-making (.79). The lowest areas were: (a) security in their job (.57), (b) staff support and facilities (.62), and (c) teacher and staff communication (.67). Ironically, student achievement or classroom pedagogy was not ranked by the participants as a factor for principal satisfaction (Friedman et al., 2008).

In order to judge the quality of instructional delivery, critics would argue that principals and assistant principals must be able to determine the amount of content knowledge as well as pedagogy that teachers have in order to make accurate assessments regarding their effectiveness in the classroom (Murnane, Singer, Willett, Kemple, & Oden, 1991). The debate over what is more important, content knowledge or pedagogical knowledge has historically plagued school leaders when evaluating staff (Shulman, 1987). According to Finn (1999), teacher preparation programs have overemphasized

pedagogical strategies as opposed to focusing on the mastery of subject area content. Subsequently, skeptics have disputed the notion that anyone can be an effective teacher simply by learning to implement scholastic practices, and they have also posited that this mentality has created a climate in present day society that anyone can be a teacher. As a result of urgent needs in many school districts, out-of-field teachers have been employed with the notion that novice educators will learn the pedagogical skills through internal district programs. This has led to an overrepresentation of novice teachers in low performing schools (Finn, 1999).

Due to the continuing debates of theorists and practitioners about the assessment of instructional quality, researchers have investigated the mental models of school leaders regarding high quality instruction. In order to understand the factors that principals believe constitute teacher effectiveness, Torff and Sessions (2005) conducted a perceptual survey of school leaders. The randomly selected sample of 300 principals from various school districts within the state of New York was comprised of 150 school principals from low performing schools and 150 from high performing schools. The principals expressed their perceptions regarding threats to teacher effectiveness based on five different areas of measurement identified by comparing 20 teacher evaluation guides across the state. Four of the areas were operationally defined and aligned with the broader concept of pedagogical knowledge; one was aligned with content knowledge. The survey yielded a return rate of 81%. According to the results from the study, all four dimensions for pedagogical knowledge were rated significantly higher by the principals as measures of teacher effectiveness than the lone area focused on content knowledge.

Interestingly, the domain that measured lesson planning was rated lower by the principals than: (a) rapport with students, (b) lesson implementation skills, and (c) classroom management. These three areas were unable to be ranked due to their very close levels of statistical significance. Therefore, the researchers concluded that the perceptions of the 300 principals overwhelmingly supported the philosophy that pedagogical knowledge was a more powerful predictor of teacher effectiveness than content knowledge (Torff & Sessions, 2005).

As indicated in the Torff and Sessions 2005 study, pedagogical practice was the preference of the principals. Because most standardized assessments have been used to determine proficiency in reading and mathematics, understanding the insight of school leaders regarding these content areas must be addressed. Medley and Coker (1987) conducted a study using a sample of principals and teachers from a southeastern portion of the United States. All teachers included in the study taught reading or mathematics and had applicable standardized pretests and posttests for their content areas of instruction. The most alarming conclusion from the study indicated the low accuracy of the judgment of the building principal when evaluating the effectiveness of the teachers. The findings also revealed that the majority of principals did not possess the skills required to assess effective instruction. Further results indicated that the perceptions of the principals were generally formed by their individual concepts of what constituted good teaching in the classroom and was solidified by their personal experiences in conducting both informal and formal observations. Standards for instruction and subject knowledge were insignificant when assessing effectiveness (Medley & Coker, 1987).

Internationally, the perceptions of school leaders regarding the purpose of conducting classroom observations has also yielded interesting findings. In a study conducted by Lam (2001), 2,400 educators which included both classroom teachers and administrators, revealed their perceptions of classroom observations. Results were calibrated for teachers' and principals' insights. Teachers indicated that they believed the primary purpose for classroom observations was to conduct appraisals; however, the majority of the instructors expressed a desire that classroom observations be utilized to target areas for specific professional learning needs. Interestingly, school principals viewed their roles of conducting classroom observation not as a primary source for evaluations but as an opportunity to determine patterns of instruction that would delineate areas for universal faculty professional learning. Finally, teachers overwhelmingly expressed a desire for their peers to be given opportunities to observe them and provide feedback regarding their instruction rather than only school principals and department heads. The instructors further stressed that peer teachers observing other educators would be more valid as well as permit the sharing of best practices (Lam, 2001).

In another study, conducted by Jacob and Lefgren (2007), principals were very successful at identifying those educators attaining the greatest as well as the least learning gains as measured by standardized assessments. However, school leaders were less accurate with the 60-80% of teachers in the middle. According to Predergast and Topel (1993), the subjective measures for evaluating teachers often leads to consensus building between school leaders and staff as to what constitutes the components of teacher effectiveness. Heneman (1986) had noted earlier that there was a relatively weak

relationship between evaluations that were subjective for teachers and the actual objective performance as indicated by the student results on standardized assessment.

School leaders receive indicators as to teacher performance throughout a school year. These data are derived from (a) formal and informal observations, (b) reports from parents including student assignment or reassignment requests, and (c) standardized achievement scores. Jacob and Lefgren (2007) expressed the belief that school leaders decipher these signals differently, and these variances in levels of expertise among administrators can impact the accuracy of their judgments regarding teacher effectiveness. In another study, only 70% of the school leaders surveyed reported that classroom observations were a favorable practice for increasing teacher effectiveness (Gordon et al., 2001). Likewise, principals reported that value-added measures, as a source of data, produced challenges for teacher effectiveness unless their students were truly randomly assigned and longitudinal assessment results were available for the same content areas.

According to Jacob and Lefgren, the correlation of value-added scores for reading and mathematics teachers and principals' perceptions of teacher effectiveness were .29 and .32 respectively. Consequently, in Jacob and Lefgren's study, the views of principals regarding what constituted teacher effectiveness were significantly higher as compared to the actual value-added scores earned by the same teachers. Furthermore, the value-added measures were a better predictor of student achievement than the perceptions of the principals (Jacob & Lefgren, 2007).

Florida and Large Urban School District (LUSD)

Policy makers have supported the view that the inconsistencies between earned value-added scores and the perceptions of school leaders regarding teacher performance in the state of Florida served as a rationale behind the Student Success Act. In the LUSD, like most school districts in the state of Florida, there has been a significant discrepancy in the value-added measures assigned to teachers and the actual student performance on the FCAT. In addition, instructional practice scores awarded by school administrators through formal and informal observations have been highly inflated when compared to the student results on the same state assessment (Large Urban School District, 2013b). Because legislation has dictated that there must be a value-added element included for each specific course beginning in the 2014-2015 school year, central office staffs were feverishly racing against time to create valid and reliable end-of-course examinations. Simultaneously, district leaders were providing inter-rater reliability trainings for all school administrators in order to increase their capacity to conduct meaningful classroom observations (Large Urban School District, 2013a).

From 2010 through 2012, school districts in the state of Florida were given latitude as to how to implement the formula (Florida State Statute 1012.34, 2012). As state statute demanded accountability through Florida's value-added model, school district policies were destined to influence the high stakes decision-making of school leaders. The analysis of how principals perceived teacher effectiveness was more crucial than ever before. According to Kelly (2004), value-added models are only valid for informing supervisors of teacher impact if the similar formative assessments are

administered to the applicable students during previous years. This has brought some difficult challenges to secondary level school leaders. For example, in specific content areas where subject matter expertise is required, such as secondary mathematics and science courses, the validity of current value-added measures may alter the viewpoints of principals and assistant principals due to a lack of comparative student data from previous years (Kupermintz, 2003). In LUSD, those teachers who were directly responsible for the instruction of reading or mathematics earned their value-added scores for the students that they instructed during both the second and third full-time-equivalent (FTE) survey periods. However, teachers outside of these areas of instruction received a school-wide average in either the area of reading or mathematics. Consequently, many teachers earned value-added scores for areas unrelated to their actual field of instruction. This was particularly true at the secondary level (Large Urban School District, 2013b). With the implementation of the Student Success Act in 2014-2015, teachers were expected to receive a value-added measure for the actual performance in their content area of instruction with their students (Florida State Statute 1012.34, 2012).

Table 4 presents the numbers of teachers assigned to categories of performance in Large Urban School District for the 2011-2012 school year as indicated by (a) value-added measures, (b) instructional practice scores, and (c) summative evaluations (Large Urban School District, 2013b). Based on an inspection of the raw numbers, it could be assumed that administrators were being more critical during their observations than the valued-added scores indicated. This was not the case. Of the total observations, 85% were at the Applying (Effective) or Innovating (Highly Effective) levels for instructional

practice observations conducted by school administrators (Large Urban School District, 2013c). Also of note is that under current practice at the time, a majority of teachers received value-added measures based on school-wide averages in the area of reading or mathematics only. As further evidence to support the implementation of course specific growth models, one must consider the district assessment summaries for reading as based on the FCAT. In Grade 3, the district had an average of 58 % students earn proficiency levels of 3 or higher, and 99.21% of teachers were rated on their summative evaluations as Effective or Highly Effective. Likewise, for Grade 8, 56 % of students earned proficiency scores; in Grade 10, however, only 52% of students were on grade level (Florida Department of Education, 2014a). Theoretically, the implementation of the Student Success Act should align the student performance with teacher effectiveness.

Table 4

Frequencies and Percentages: 2011-12 LUSD Teachers Assigned to Performance Categories (N=12, 558)

Categories	Value-added Measure		Instructional Practice Score		Summative Evaluation	
	N	%	N	%	N	%
Highly Effective	460	3.60	154	.12	165	1.31
Effective	11, 905	93.27	11,647	92.72	12,295	97.90
Needs Improvement	139	1.09	471	3.75	64	.50
Developing	43	.34	268	2.13	32	.25
Unsatisfactory	11	.0008	18	.001	1	.00007

Source. Large Urban School District, 2013c

Principals' Decision-Making Using the Data

School systems in the 21st century have been obligated to hire and retain effective educators even though most educational institutions struggle with operationally defining the phrase, highly qualified (Cornett & Bailey, 2003). This section of the review focuses on how school administrators have been reported to utilize data from classroom observations, value-added measures, and achievement scores to make personnel decisions.

The Principal as the Leader for Improving Teacher Quality

Effective principals understand that their primary role as the instructional leaders of their schools is to improve the teaching of their instructional staff. This shift in leadership has stemmed from studies showing that by augmenting the performance of the educators within the schools, increases also occur in students' achievement scores (Wallace Foundation, 2013). This transformation in focus has required that school leaders become skillful at interpreting and conveying pertinent data to classroom teachers in order to adjust instruction. This task is very complex, as researchers have revealed that there is a wide spectrum of expertise among teachers within each school. According to Darling-Hammond (2000), new teachers are less effective than seasoned veterans at increasing student learning. However, experienced teachers have often been found to be more resistant to changing their practice of instruction to accommodate the changing needs of their students. The challenge, therefore, is for effective school leaders to determine how best to use their school data.

Savvy school leaders create a working environment that promotes autonomy among teachers while concurrently creating a school culture that embeds action research as a venue for improving instruction. Donaldson (1993) observed that through action research, administrators are able to assist teachers in using data to focus on reflective questions regarding student achievement, instructional practice, student management, and other pertinent classroom behaviors. Recent emphasis given to intensive analysis of the instructional practice of teachers and the value they add over time regarding students' learning gains, has provided validity to the profession (Watkins, 2005).

Data is only effective if it promotes a change in practice. Because teaching is both an art and a science, school leaders must use the data to drive the discussion with teachers (Marzano, 2007). Effective school leaders are masterful at holding courageous conversations with teachers who struggle with instruction.

According to Jackson (2008), administrators can have four different types of conversations with teachers: (a) reflecting, (b) facilitating, (c) coaching, or (d) directing. In reflecting conversations, school leaders can guide teachers through a process that affords them the opportunity to discover how their beliefs and behaviors impact instruction in the classroom. Through facilitative conversation, school administrators direct discussions towards the instructional goals. In the coaching phase, the principal or assistant principal begins to specifically target areas in which instruction needs to improve. During this interaction, the administrator suggests strategies, corrections to existing approaches, and emphasizes professional growth. Finally, in directive conversations, the school leader uses the data in order to provide clear and specific

instructions as well as consequences if not followed. After directive conversations, school administrators should expect immediate action on the parts of teachers to alter their instructional approach (Jackson, 2008).

According to a Wallace Foundation report (Shelton, 2010), robust data systems assist school leaders in connecting the evidence obtained from classroom observations with teachers' student data. A thorough evaluation system will accomplish five tasks. First, effective evaluation systems will provide continuous feedback to principals and assistant principals in order to record the individual educators' progress towards practicing the strategy and increasing their content knowledge. Second, useful evaluation systems will assist teachers in satisfying requirements for professional licensure. Next, valid effective evaluation systems must diagnosis the weaknesses of individual teachers and identify the required professional learning and support needed to become highly effective. In addition, superior evaluation systems will provide feedback to universities, colleges, and state departments in order to provide direction for preparatory programs. Finally, summative evaluation systems have to provide the necessary information for state and federal accountability (Shelton, 2010).

Regardless of systemic controls, such as powerful data technologies supported by valid and reliable metrics, the humanized factor of leniency has far too often prevented school administrators from recording objective data during classroom observations. Weisberg et al. (2009) noted that 98% of teachers were rated as satisfactory or above when determined by classroom observations. However, as instructional frameworks are coupled with significant administrative training, there have been examples of linking the

classroom observation of teacher practices with student achievement data. In a study conducted by Kane, Taylor, Tyler, and Wooten (2010) of Cincinnati's evaluation system, evidence supported the relationship between the scores earned during classroom observations and the student achievement scores on the annual standardized assessment completed by students. For school leaders, this information was extremely valuable even if the majority of the teachers earned a rating of satisfactory or higher. Principals were able to differentiate the levels of satisfactory of their teachers and determine individual improvement efforts for members of their staff. Furthermore, observing the instructional delivery of individual teachers afforded administrators opportunities to focus on patterns of pedagogical practice in order to foster large scale growth opportunities (Kane et al., 2010).

According to Florida State Statute 1012.28 (2012):

Each school principal is responsible for the performance of all personnel employed by the district school board and assigned to the school to which the principal is assigned. The school principal shall faithfully and effectively apply the personnel assessment system approved by the school board pursuant 1012.34.

The principal is responsible for the evaluation system and may assign evaluation responsibilities to assistant principals assigned to the school building. (p. 1)

Florida State Statute 1012.34 (2012) has required districts to provide observation instruments and criteria for continuous quality improvement of professional skills of school personnel. Furthermore, performance evaluation results must be utilized in determining professional learning needs for teachers (Florida State Statute 1012.34,

2012). Therefore, data driven instruction is the goal of professional learning (Pella, 2012). According to Marzano (2004), the professional learning needs of the teachers must be obtained from a multitude of data sources including both formative and summative assessments of student achievement as well as classroom observations. This notion has been supported by large urban school districts around the country. In 1998 in San Diego, the school district prioritized the professional learning as extremely low. Less than 1% of the district's budget was allocated to furthering the knowledge of instructors and their leaders. Leaders realized that in order to promote student achievement, the district would have to invest heavily in developing the work force. By 2005, approximately 6.5% of the entire budget was being devoted to increasing the performance of the instructors within the school district. However, like many districts, professional learning was originally prescribed for all teachers equally. As time progressed, individual teachers became more involved and vocal regarding the universal approach. Consequently, the district began to tailor learning opportunities to individual employees. A commitment from the school district differentiated professional learning for the individual teachers and principals based on the actual student performances within their classrooms and schools. Overall, 56% of the English language teachers surveyed viewed their professional learning opportunities as being aligned with their students' needs, and 55% indicated that these experiences were associated with their personalized professional goals (Quick, Holtzman, & Chaney, 2009)

At the time of the present study, LUSD teachers were required to incorporate the Deliberate Practice model in order to improve their practice, and principals and assistant

principals of the LUSD were required to attend eight days of training in order to evaluate the success of their teachers (Large Urban School District, 2011). With Deliberate Practice, classroom teachers, with the assistance of their administrators, have been able to select various strategies on which to focus as opposed to attempting to master an overwhelming array of instructional practices. At the beginning of a school year, teachers select one routine strategy, one content strategy, and one strategy enacted “on the spot.” For instance, an individual educator has the opportunity to choose a strategy for communicating clear learning goals from the general category of routines, a strategy for previewing new information from the content category, and a strategy for using academic games to engage students from the category of strategies enacted on the spot. When teachers are able to choose their own strategies in order to improve, a sense of ownership for the process becomes a routine practice (Marzano, 2010).

The principals and assistant principals of the LUSD have utilized the iObservation instrument when observing classroom teachers for instructional practice. The Marzano evaluation model has several constructs worth noting: First, teachers can increase their expertise with effective strategies from year to year, and this can produce significant gains in student learning. Second, a common language exists in order to communicate between school leaders and classroom teachers. Furthermore, this common language reflects the complexity of teaching. Finally, the school leader provides focused feedback and targeted practice that utilize the common language for professional growth (Large Urban School District, 2011). The evaluation contains four domains as follows: (a) Domain 1: Classroom Strategies and Behaviors, (b) Domain 2: Preparing and Planning,

(c) Domain 3: Reflecting on Teaching, and (d) Domain 4: Collegiality and Professionalism. This instrument contains 10 design questions, each containing key elements for the school administrator to rate the teacher based on one of five scales. The five scales are as follows: (a) *Not Using*, (b) *Beginning*, (c) *Developing*, (d) *Applying*, and (e) *Innovating*. Each scale is equated to a numeric value of 0-4 respectively. These ratings are determined by evidence observed by the school leader. The purpose of determining the teacher's proficiency in using these elements is to promote conversation between teachers and administrators for the purpose of improving classroom instruction. Following is a description of the process employed by school leaders during conversations in order to increase teachers' ability to move to a higher level on the scale (Marzano, 2007).

When school leaders observe teachers as *Not Using*, they should enlighten the teachers to specific research along with applicable strategies to use in the classroom. When teachers attempt these strategies, they may be augmented to the *Beginning* level. During this phase, teachers are omitting or using the strategy with errors. The school leader's role is to identify the errors and make teachers aware. Once teachers master the strategies and perform without errors, they have moved to the *Developing* level. In this phase, teachers are performing the strategies correctly but are not monitoring the students effectively. When teachers are observed at the *Developing* level, the school leader provides suggestions on how to increase their monitoring of the student responses. Once they master the student monitoring of responses, they have augmented and move to the *Applying* level. Finally, the principal or assistant principal provides feedback to teachers

that stresses providing macro strategy for struggling students. This ensures that all students are receiving and able to apply their new knowledge. This phase is called *Innovating* (Marzano & Simms, 2013).

Teacher Renewal

Once school leaders have the data and identify areas of growth required for staff, it is necessary to analyze how the data are utilized in order to impact the management of human capital (Donaldson, 2009). Constituents of elected officials have argued that school leaders must prioritize goals of the school district with the practices of the classroom teacher. In addition, many citizens have demanded that retention and compensation of teachers be linked to student performance (Tucker, 2001). In addition, schools that have higher student achievement have leaders who hire, train, assign, and retain differently than those with lower student performance. Furthermore, because it usually the school principal who makes these decisions, having an instructional leader who is talented in the area of human capital management often produces the desired effects on student learning (Loeb, Kalogrides, & Beteille, 2012).

As the pressure to supply high quality teachers increases, the survival of school leaders will be increasingly dependent upon their skills at selecting staff who will continuously improve the performance of their students. Generally, when asked, the majority of school principals support the notion that the hiring of teachers is one of the most important functions they have as school leaders. According to Mertz (2010), principals believe it is an opportunity to improve the instruction by adding talent to the

faculty. However, what is stated often contradicts what is practiced. In a study conducted by Mertz (2010), 57 principals representing 23 school districts in the southeastern United States were interviewed. Of the 57 participants, 33 were secondary principals, 22 were elementary principals, and 3 were private school principals. Over three years, participants were interviewed individually about the following components for selecting instructors at their schools: (a) how they conducted their interviews, (b) who was involved in the decision process, (c) who made the decision to hire, and (d) to what extent they felt they had all the information necessary to make a selection. Interestingly, both elementary and secondary principals reported similar responses to the research team. Ironically, the study revealed that most principals did not use the teacher selection process as a vehicle to improve teacher instruction or student learning. Principals based their selection on instinct based on their first impression of the applicants. Many boasted that they knew within the first two minutes of the interview as to whether or not they would hire the teacher. Of the 57 principals, 40 reported use of a committee to interview in order to determine a good fit for the school. Some indicated that this relieved them of some pressure, i.e., solely owning the decision. Furthermore, only one principal of 57 reported using his vision and mission for the school as a basis for selecting qualified candidates. The researchers concluded that the principals in this study did not utilize the teacher selection process as a vehicle to influence the instructional direction of their schools by employing teachers who had known attitudes, competencies, and knowledge associated with traits for improving student performance (Mertz, 2010).

Once teachers are employed, school leaders are obligated to make decisions that will impact the educator's continued service within the school and district. There is little argument that contract renewal has been a topic of much debate over the years (Alexander & Alexander, 2012). Teacher unions have challenged policy makers with regard to tenure, indicating that it is a right of educators in order to ensure that individual teachers are not victimized by arbitrary terminations (Winters, 2012). According to a study conducted by Tucker (2001), ineffective tenured teachers account for approximately 5% to 15% of the teachers annually across the nation, but less than 1% of tenured teachers are dismissed. A pitfall of the tenure system, according to Tucker, is that ineffective teachers as well as those that contribute towards the betterment of their students, are protected. Many experts have posited that students are in jeopardy when a system functions to protect those who are ineffective. Other opponents of tenure practices have noted that teachers are often tenured early in their careers, and this reward has very little to do with their actual performance in the classroom (Winters, 2012).

With the importance given to accountability under legislation from Race to the Top, one would assume the dismissal of ineffective teachers would be based largely on student performance (U.S. Department of Education, 2013). In a study conducted by Nixon, Packard, and Dam (2011), 544 principals from the southeastern United States participated in a survey regarding the dismissal of ineffective teachers. Questions pertained to the following categories: (a) absenteeism and tardiness, (b) classroom management, (c) ethical violations, (d) incompetence, (e) professional demeanor, (f) insubordination, and (g) lack of student achievement. Each of these categories was rated

on a scale where 1 = most unlikely to 7 = most likely. Results indicated that the majority of non-renewals of teachers by principals was due to specific ethical violations. In fact, 325 respondents assigned a “7” rating to specific ethical violations as the most important indicator for terminating the employment of a teacher. The researchers hypothesized that because accountability was heavily emphasized in recent legislation, principals would rank “lack of student achievement” as a key component for decision making. Of the responding school administrators, elementary principals ranked this criterion as more important than secondary principals. However, neither group rated it more than moderately high. With achievement scores often being lower in urban settings, these secondary principals ranked lack of student achievement as more important than their peers in suburban districts. To the shock of the researchers, of the 544 participants, only 24 principals ranked lack of student achievement as a most likely criterion for teacher dismissal. Perhaps more surprising was that 210 principals indicated that lack of student achievement was rated as unlikely, very unlikely, or most unlikely as a reason for teacher dismissal (Nixon et al., 2011).

Many states have recognized the practice of principals in being tolerant of ineffective teachers and have compensated by enacting strict statutes to expose and sanction ineffective teachers and principals. By mandating the use of statistical evaluation tools that remove the human factor, ineffective instructors are revealed and contracts threatened. These states endorse the concept that there is a statistical measure that can be assigned to teachers in order to measure the value that they contribute to their students’ annual learning growth. Value-added models can assist school leaders in

making more informed decisions about who should receive tenure or contract renewal.

Colorado, Tennessee, and New Jersey passed legislation dictating that teachers receiving below satisfactory performance ratings for two consecutive years will lose their tenure and be dismissed from service (Winters, 2012).

In a study conducted in Florida, a sample of second year pre-tenured teachers participated in a three-year study. Winters (2012) evaluated the relationship between the value-added measure and the student achievement scores on the FCAT for all teachers during their fifth year of teaching. He determined that the VAM scores of the pre-tenured teachers provided relevant information regarding the future ability of the teachers. He also concluded that a VAM based tenured policy would have removed these teachers that performed worse than their peers later in their career. Therefore, according to Winters, “The results tell tenure reformers that they should consider the number and type of teachers likely to be denied tenure or removed from the classroom under their proposed policies” (p. 7).

For the LUSD, the summative evaluation for teachers derived data from classroom observations, value-added scores from the FCAT learning gains, translates categories for continued employment. Categories are comprehensive categorizations that have been developed for instructional personnel based on their experience and performance. Consensus for these categories was achieved through collective bargaining with the teachers union and approved by the department of education (Florida State Statute 1012.34, 2012).

Category 1 teachers are within their first two years of the profession. These teachers do not receive value-added scores as they have not had previous students. Teachers may earn instructional practice scores from classroom observations. These teachers are placed on an annual contract and can be dismissed by the school principal due to poor performance or budgetary reasons. (Large Urban School District, 2013a).

Category 2A teachers are in at least their fourth year and earn a value-added score connected to their students' learning gains on FCAT along with an instructional practice score (Large Urban School District, 2011). Teachers who reached the end of their third year prior to July 1, 2011, could have been granted tenure by their school districts. If they did not, they receive annual contracts for continued service (Florida State Statute 1008.22, 2012). Category 2B are experienced educators with at least three years of teaching experience with one of the following applications: (a) newly hired teachers to the district, (b) teachers who have been assigned to instruct a new subject that is different from their previous assignment, (c) teachers employed at a school with a different student population from previous year, or (d) teachers who earned between a 2.0 and a 2.4 for an instructional practice score for the previous year (Large Urban School District, 2013a).

For the LUSD, summative data has been used to put ineffective teachers on notice that improvement must occur in order to retain their positions. This domain, Category 3, is for teachers who have been rated ineffective in the classroom either through classroom observations or earned value-added scores that delineate low student growth. Category 3 teachers receive additional support from their administration and regular feedback on areas needing improvement. School principals and assistant principals monitor the

teachers with a Professional Improvement Plan (PIP) that includes additional observations to monitor for progress. Once the PIP is completed, the principal reassigns the teacher to the original classification. However, if Category 3 teachers are unable to meet the requirements of the improvement plan, they may receive an overall Needs Improvement or Unsatisfactory rating on the final evaluation. Finally, there is one more group for teachers: Category 4. Teachers are placed into Category 4 because there is a lack of sufficient data in all four domains for instructional practice to be evaluated equitably due to one of the following conditions. This can be due to an extended absence or starting their current assignment after February 15th of the school year (Large Urban School District, 2011).

Summary

This review of literature illustrates the impact of the federal mandates under Race to the Top that hold school leaders and classroom teachers accountable for the performance of their students (U.S. Department of Education, 2013). Furthermore, the federal requirements have been tightly coupled with language in Florida state statute that directs all school districts to implement an instructional evaluation system based on the value teachers add to their students' learning growth (Florida State Statute 1012.34, 2012). Classroom observations, value-added models, and student growth have transformed the traditional roles of school leaders from building managers to instructional leaders (Ravitch, 2010).

Although classroom observations have been a traditional practice of school leaders since the beginning of the 20th century, what principals are looking for has changed. As pedagogy has gained increased focus, scholars such as Marshall (2010), believe that school leaders should conduct much more frequent observations and analyze different instructional strategies than in previous years. Most districts have developed instructional frameworks that guide the observations of school leaders through various domains in order to allow them to provide specific feedback to teachers (Marzano, 2007). In addition to providing data tools to monitor instructional delivery, principals have also begun to have different conversations with teachers than they did in the past. In districts like the LUSD, teachers have been encouraged to take an active role in their own professional growth. Feedback has been modified to provide for interactive conversations initiated by the school leader in order to assist teachers with improving their craft. This deliberate practice has been designed to help teachers refine very specific elements of instruction for the purpose of improving their students' performance (Marzano, 2007).

Much of the public appeal regarding value-added models can be linked to the original work of Sanders and Horn (1994). In their work, they attempted to neutralize factors affecting student outcomes, including such issues as (a) mobility of students; (b) modes of teaching; and (c) altering teacher assignments. Their underlying conclusion was that effective teachers tend to be so regardless of the proficiency levels of their students (Ravitch, 2010). This notion of excuse-free accountability was the premise behind Florida's work. As such, Florida took the responsibility of adopting a value-

added model very seriously. The analysis and development of the covariate adjustment model used within the state received the attention of both legislators and the Florida Department of Education (2014). Leading members from various stakeholder groups analyzed a host of models before making their recommendation to the Commissioner. The initial model was built to satisfy the requirement that each teacher contributed to learning gains in reading or mathematics. However, by 2014-2015, each course from the Florida Course Code Directory was to have an assessment provided displaying the valued added by the teacher.

Although value added is a fairly new concept for Florida educators, learning gains is not. This term has been used to define comparable individual student performance annually on the Florida Comprehensive Assessment Test (Florida State Statute 1008.22, 2012). Though learning gains are specific to Florida, the notion of student growth as measured by standardized assessment has become a national expectation for 21st century school leaders and classroom teachers. The alignment of the FCAT learning gains and the value-added for each teacher is the reality of being a classroom teacher in the state of Florida (Florida State Statute 1012.34, 2012). For Florida educators, understanding the complexities of the standardized assessments items and their relationship with higher order thinking skills are imperative in order to properly align instruction with assignments (Hess et al., 2009). Consequently, educational leaders have the responsibility to give precedence to the required shifts that are necessary in order to produce high quality instruction through research based instruction (Hashey & Connors,

2003). As the quality of teaching improves, the established growth in student learning is inevitable (Hattie, 2009).

Even with the changes in expectations, school leaders have struggled with altering their expectations and perceptions of what constitutes effective instruction. Researchers have revealed that principals are more concerned with pedagogical practice than actual content knowledge (Torff & Sessions, 2005). According to Jacob and Lefgren (2007), although principals are usually accurate in judging the most and least effective teachers, most principals still struggle with identifying the needs of teachers in the middle of the effective spectrum. In the LUSD, the target district in the present study, the majority of school leaders have a history of identifying the majority of the teachers as Effective or Highly Effective. However, student achievement scores have not yielded corresponding results. To further complicate the issue, despite legislation and applicable sanctions, according to Mertz (2010), school leaders have persisted in selecting teachers based on their first impressions regarding individuals' fit to the school rather than their instructional effectiveness. Related to this topic, student achievement has been ranked very low as a reason for voiding a professional contract or revoking tenure (Nixon et al., 2011).

This literature review has shown that effective instruction is deemed so based on the data collected by administrators that ultimately proves it. School leaders have been required to become scholarly practitioners who must justify their actions for awarding praise as well as sanctions. Data systems allow school districts to collect volumes of information on the value-added by each teacher that directly corresponds with the

performance of their students. High stakes decision making, including continued employment decisions, has become the norm for school leaders in their use of summative information. The traditional right of tenure is no longer, and in the state of Florida for the educators hired after July 1, 2011, it is an urban legend. Teachers may move up and down a classification annually, and dismissal has come to be a reality of poor performance (Florida State Statute, 1012.34, 2012).

Chapter 3 contains a discussion of the methods and procedures used to conduct the study. The methodology is explained for the analysis used within the LUSD regarding data from 2012-2013 pertaining to: (a) the instructional practice scores assigned by school principals and assistant principals for instruction to secondary reading teachers, (b) the value-added measures assigned to the secondary reading teachers, and (c) the learning gains earned by their students.

CHAPTER 3 METHODOLOGY

Introduction

During the 2012-2013 school year, under the requirements of the Race to the Top grant, the state of Florida and all local educational agencies were required to ensure that 60% of teachers' evaluations would include the measures earned on the instructional practices section of the observation instrument adopted by the Large Urban School District (Florida State Statute, 1012.34, 2012). The remaining 40% of teachers' evaluations were to be constructed using a measure of growth in student learning. For teachers of reading and mathematics in Grades 4-10, the state of Florida calculated value-added scores to be used for this portion of the evaluation.

Secondary principals and assistant principals for instruction were given the responsibility of gauging the effectiveness of teachers based on evidence observed during classroom instruction. Teachers were rated based on their quality of the implementation of the learning goal as rated against an established system of evaluation scales (Marzano, 2007). As of May 1, 2013, all instructional personnel were required to receive one of the following preliminary ratings: (a) Highly Effective (b) Effective, (c) Needs Improvement (d) Developing, or (e) Unsatisfactory (Large Urban School District, 2011). The Needs Improvement and Developing ratings were equivalent, though the Developing rating was only available to teachers in their first three years of service in the state of Florida.

In late July of 2013, the state of Florida assigned each teacher a value-added score in accordance with Florida State Statute 1012.34 (2012) based on a growth model for

those students who were in attendance at selected points in the school year (Full-Time-Equivalent Surveys 2 and 3). The state of Florida first produced value-added scores for teachers in the summer of 2011 with the assistance of its statistical partners at the American Institutes for Research. For this study, the value-added score assigned to teachers for the 2012-2013 school year were designed to reflect the learning gains for students based on FCAT results that were assigned to specific teachers. The students associated with teachers for the LUSD were attached to teachers at either Survey 2 or Survey 3 so long as the students were present at the same school for both Survey 2 and Survey 3. This value-added score was used for the remaining 40% of teachers' overall evaluations (Florida Department of Education, 2014a).

The LUSD was comprised of (a) 994 site-based and district administrative personnel, (b) 13,196 instructional staff, and (c) 189,347 students. Administrative racial distribution consisted of: (a) 56% White, (b) 26% Black, (c) 13% Hispanic, and (d) 5% Other. Student racial distribution was: 30.3% White, 26.9% Black, 35.4% Hispanic, and 7.4% Other. The percentage of students eligible for free and reduced lunch was 66% (Large Urban School District, 2014).

This research was conducted in response to Florida state statute 1012.34 (2012), which mandated specific evaluation criteria that school administrators must use when assessing teacher effectiveness. This chapter consists of five sections. The first section describes the purpose of the study and the research questions. The second section reveals information regarding the participants of the study. The third section elaborates on the development of the survey used to collect the data from the participants. The fourth

section defines the data collection procedures. Finally, the data analysis is explained in the fifth section.

Purpose of the Study

The purpose of this research study was to determine to what extent a relationship existed between the instructional practices portion of the teachers' summative evaluations conducted by secondary principals and assistant principals for instruction, the assigned valued-added score based on student growth in the area of reading for Grades 6 through 8 and 9 through 12, and the learning gains earned by teachers.

Research Questions

The following research questions were formulated as the focus of this study.

1. To what extent was there a relationship among each administrator's evaluation of teachers' instructional practices, the value-added measure, and the learning gains assigned to teachers from learning gains as measured by Florida Comprehensive Achievement Test (FCAT) Reading for the following grades: (a) 6 through 8 and (b) 9 through 12 in an urban school district for the school year 2012-2013?

H_0 . There is no relationship between instructional practice scores attained through administrators' observation, the value-added measures, and the learning gains assigned to teachers based on the learning gains of their

students in reading as determined by the FCAT in Grades 6 through 8 or 9 through 12.

H_A. There is a relationship among the instructional practice scores attained through administrators' observations, the value-added measures assigned to teachers, and the learning gains of their students in reading as determined by the FCAT in Grades 6 through 8 or 9 through 12.

2. What factors do middle and high school principals and assistant principals believe contribute to the relationships among the instructional practice ratings, the value-added measures, and the learning gains?
3. To what extent do principals report using the instructional practices score, learning gains, or VAM scores to make personnel or instructional decisions?

Participants

The study consisted of all secondary school principals and assistant principals for instruction in the LUSD for the 2013-2014 school year. To participate, administrators were required to have observed secondary school reading teachers during the 2012-2013 school year. The sample studied consisted of 138 school based administrators (65 principals and 73 assistant principals for instruction). Specifically, the sample consisted of 44 principals and 42 assistant principals for instruction at the middle school level and 21 principals and 31 assistant principals for instruction at the high school level. Of the middle school principals, 60% were female, and 40% were male. Of the middle school assistant principals for instruction, 73% were female, and 27% were male.

At the high school level, 47% of the principals were female, and 53% were male. Of the high school assistant principals for instruction, 71% were female and 29% were male. Table 5 reflects the racial composition of principals and assistant principals for instruction at the middle and high school levels.

Table 5

Racial Composition of Participants: Principals and Assistant Principals for Instruction (API)

Race	Middle School		High School	
	Principal	Assistant Principal	Principal	Assistant Principal
White	60%	60%	88%	47%
Black	20%	34%	11%	21%
Hispanic	15%	1%	1%	26%
Other	5%	5%	-	6%

Source. Large Urban School District, 2014

The sample of 138 secondary principals and assistant principals for instruction were surveyed regarding (a) instructional practice observations, (b) value-added measures, and (c) learning gains for reading at their schools. The purpose was to determine if a significant difference existed between the three measures for the LUSD being studied. Results of this study had generalizability for school districts within the state of Florida that used the same Marzano iObservation instrument. Although other school districts within the state may have chosen a different observation tool and applied the value-added model differently, Florida State Statute 1012.34 (2012) required all school districts to use a classroom observation instrument aligned with the standards to

be taught and to utilize the assigned value-added scores for at least 40% of teachers' final evaluations. In addition, the learning gains associated with all teachers across the state were derived from the Florida Comprehensive Assessment Test (FCAT 2.0) (Florida State Statute 1008.22, 2012). Though the FCAT 2.0 was scheduled to change to another format, the new assessment instrument was intended to remain standardized across the state of Florida to provide the same information to principals about their teachers and students. This was structured to allow the state to construct value-added ratings of teachers without interruption during the transition to new statewide assessments in 2014-15 (Florida State Statute 1012.34, 2012).

Instrumentation

Each participant received an electronic communication (Appendix C) from the researcher prior to receiving the perceptual survey. The Instructional Practice, Value-added Measure, Secondary Reading Learning Gains Survey (Appendix D) was developed by the researcher for the purpose of answering Research Questions 2 and 3. The researcher developed the 39-item survey after conversations with school district administrators involved with the three components of this study and university faculty. The content of the survey was reviewed by experts at the school district and university for content validity. Revisions were made based on the feedback the researcher received.

The researcher designed the survey in an electronic format to be delivered to the participants individually. The Internet-based survey was selected by the researcher due to the high degree of functionality of the instrument as well as ease of completion by the

participants. The format of the survey permitted the researcher to secure data quickly and easily. The researcher piloted the survey with select district administrators who previously held school site secondary administrative positions and were familiar with the three components of the study. The rate of return for the pilot survey was 100%. Feedback was provided to the researcher from the pilot survey participants. Edits were made accordingly.

The survey consisted of five sections: Section 1 (items 1-8) was used to elicit demographic information regarding the school and the administrator. Section 2 (items 9-18) pertained to the instructional practice portion of the evaluation for secondary reading teachers. Section 3 (items 19-28) addressed value-added measures assigned for reading teachers for the 2012-2013 school year. Section 4 (items 29-37) asked participants to reflect on the reading gains for the 2012-2013 school year. A Likert-type scale ranging from 1-5 was used to record participants responses for items in Sections 2, 3, and 4. Section 5 requested open-ended responses to items 38 and 39. Item 38 asked participants to share with the researcher how they as school leaders used the results from the 2012-2013 school year with instructional practice, value-added measures, and reading learning gains to improve their schools. Item 39 requested that the participants provide the researcher with any other comments that they believed would assist in improving reading teacher effectiveness.

Data Collection

This study consisted of mixed methods of quantitative and qualitative data collection. Quantitative data were secured through school district databases for: (a) instructional practice scores, (b) value-added measures, and (c) FCAT reading learning gains in order to address research question one for the study. The researcher developed a perceptual survey to gain additional quantitative data from principals and assistant principals for instruction for research question two. In addition, qualitative data were also retrieved from the survey to respond to the third research question. The general procedures used in data collection are described in the following section followed by the specific procedures used in collection both quantitative and qualitative data.

Data Collection Procedures

Once approved by the school district designee, a communication authorizing the study was received (Appendix A). Consequently, the researcher submitted the survey along with the proposal to the Institutional Review Board (IRB) at the University of Central Florida in October 2013. In November 2013, the IRB approved the study (Appendix B).

In April 2014, a pilot survey was submitted to 10 school district administrators with secondary school principal and assistant principal for instruction experience. The survey was completed by all 10 pilot participants and feedback was secured by the researcher. Adjustments were made based on their feedback to both the instrument and the electronic formatting. After the survey was piloted, the researcher made the first

contact with the participants by sending an email to the following groups: (a) middle school principals, (b) high school principals, (c) middle school assistant principals for instruction, and (d) high school assistant principals for instruction. The email briefly discussed the purpose of the study, indicating that participants would be receiving a brief 39-item survey within a few days and praised their anticipated participation. In the time between receiving the first email and sending the actual survey, the researcher made the second contact by visiting the monthly meetings for the assistant principals for instruction and the secondary principal meetings in order to verbally explain the purpose of the study and once again request participation in completing the survey. The third contact, the actual electronic survey, was sent within 24 hours of the face-to-face meeting with the school administrators. Copies of all the correspondence with participants is included in Appendix C. (Dillman, Smyth, & Christian, 2009).

Although the researcher knew the identities of the original 138 in the sample, their responses to the survey were completely anonymous. After the survey was sent, the researcher received responses from 62 participants, a response rate of 44%. The fourth contact was made a week later via email thanking all those who completed the survey and requesting the participation of those who had not yet responded. After the fourth contact, an additional 54 surveys were completed for a total of 116 of 138 administrators responding yielding a return rate of 84%.

Quantitative Data Collection

In October 2013, the researcher met jointly with the Senior Director for Professional Development and the Senior Director for Accountability, Research, and Assessment to define the type of data to be collected to complete the study.

For Research Question 1, through consultation with the school district experts, it was determined that the instructional practice scores for reading teachers would be obtained from the iObservation instrument used by the LUSD to collect both formal and informal observations on teachers. This instrument provided principals and assistant principals information regarding a teacher's competency levels in delivering instruction based on the design questions in the Marzano framework. The database was managed by the LUSD Professional Development Department. For the purpose of this study, a report was generated to isolate the scores from the iObservation database of secondary reading teachers. This report included data for those teachers who taught the subject of reading, as defined by Florida's Course Code Directory, during both the October and February full time equivalent survey periods for the 2012- 2013 school year (Florida Department of Education, 2012). An additional report was also created from the iObservation database to capture the summative instructional practice scores for the 2012-2013 school year for applicable teachers.

For Research Question 1, the value-added measures were the metrics assigned to the individual teachers in the LUSD based on their students' growth during the 2012-2013 school year as measured on the FCAT Reading. The value-added model calculates predicted values for students in Grades 4 through 10 based on multiple student covariates

that include prior year score(s), Exceptional Student Education (ESE) status, English Language Learners (ELL) status, and daily attendance. The difference between the predicted performance of students and their actual performance is used to calculate the teacher's value-added score (Florida Department of Education, 2013d). In reviewing the intent of the research question with the school district's experts, it was determined that the value-added scores would be obtained from the Educational Data Warehouse (EDW) system utilized by the LUSD. The school district department responsible for this function was the Accountability, Research, and Assessment Department. A specialized report was generated to secure the value-added scores for secondary reading teachers for the 2012-2013 school year. Filters in the report allowed the researcher to isolate only those teachers who taught reading as a course, as defined by Florida's Course Code Directory, and who were employed by the LUSD in both October of 2012 and February of 2013 (Florida Department of Education, 2012). The data were checked by school district personnel to ensure accuracy of the two data sets. The school district has a total of 955 Reading teachers in Grades 6 through 10 and 883 of these teachers (92.5%) met the requirements for inclusion in the study.

Regarding the FCAT Reading learning gains, each teacher in the LUSD received a percentage of students earning learning gains for each of the courses they taught. This was determined by the state of Florida by comparing the students' prior performance on the FCAT with the current year's assessment results (Florida Department of Education, 2012). Learning gains are calculated in three ways: (a) an individual student may improve one or more achievement levels, (b) a student may maintain a proficient

achievement level of at least a 3 without decreasing a level, or (c) a student may demonstrate more than one year's growth when remaining a level 1 or 2. After consultation with school district experts, it was determined that these data would be retrieved with the assistance of the Accountability, Research, and Assessment Department. The data were extracted from the EDW through a school district dashboard that calculates learning gains for every reading and mathematics teacher in the school district. A report was generated to produce a total percentage of all students instructed making learning gains for each secondary reading teacher who taught the subject of reading as a course, as defined by the Florida Course Code Directory, during the October of 2012 and February of 2013.

For Research Question 2, it was determined that the secondary principals and assistant principals for instruction would have to be able to reflect on the data from instructional practice scores, value-added measures, and learning gains from the 2012-2013 school year to determine the extent that these metrics impacted their decision-making for the 2013-2014 school year.

With feedback from the school district staff, as well as from faculty experts from the university, the researcher drafted a perceptual survey (Dilman, Smyth, & Christian, 2009). To gauge the various backgrounds of the participants, Section 1 solicited demographic data from participating administrators regarding their professional experiences and formal education along with pertinent student information from their schools. The survey was designed to emphasize each of the three areas of the study: (a)

instructional practice scores, (b) value-added scores, and (c) learning gains as Sections two, three, and four respectively.

Qualitative Data Collection

Section 5 of the perceptual survey was designed to gather responses to two open-ended items from participants to respond to Research Question 3. Although no statistical test was used in this portion of the analyses, the information obtained was valuable in gaining the insights of practitioners and was designed to provide pertinent information for school district leaders. Item 38 asked respondents to reflect on how they, as instructional leaders, used the data in order to improve the academic performance at their school. Item 39 requested that participants offer suggestions for improving the effectiveness of reading teachers' instruction.

Data Analysis

The data analysis was performed to respond to the research questions which guided this study. The following descriptions of the procedures used to analyze the data have been organized by type of data and are presented for each of the three research questions. Table 6 displays the sources of data, variables, and statistics used in the data analysis to answer each research question.

Table 6

Research Questions, Sources of Data, Analysis, and Variables

Research Questions	Data Sources	Data Analysis	Variables
To what extent was there a relationship between each administrator's evaluation of teachers' instructional practices, learning gains, and the value-added measures assigned to teachers as measured by FCAT Reading for the following grade levels: (a) 6-8 and (b) 9-12 in an urban school district?	Instructional practice scores	Pearson r coefficient	Independent: Instructional practice scores
	Assigned value-added measures	Pearson r coefficient	Dependent: Value-added measures and learning gains
	Learning Gains	Pearson r coefficient	
What factors do middle and high school principals and assistant principals believe contribute to the relationships among the instructional practice rating learning gains, and value-added measures?	Instructional practice, learning gains, and value-added measure survey of principals and assistant principals	Descriptive Statistics means and Standard Deviations	Independent: Instructional practice scores Dependent: Value-added measures and learning gains
To what extent do principals report using the instructional practices score, learning gains, or VAM scores to make personnel or instructional decisions?	Open Ended question as part of the survey for principals and assistant principals	Categorize responses	Independent: Instructional practice scores Dependent: Value-added measures and learning gains

Quantitative Data Analysis

The data for instructional practice scores, value-added measures, and learning gains for reading were uploaded into SPSS version 22 for statistical analysis. Data for Research Question 1 included the instructional practice summative scores, assigned value-added scores, and learning gains for all secondary reading teachers within the LUSD (Large Urban School District, 2014). A Pearson r was calculated in order to analyze the relationship between the independent variable of instructional practice scores and the dependent variable of value-added measures. A Pearson r was also calculated in order to describe the relationship between the instructional practice scores and the mean of the learning gains for each secondary reading teacher. Steinberg's Correlation Table was used in order to determine the significance of the relationship between the independent variable and the dependent variables (Steinberg, 2011). For Research Question 2, the data collected from the 116 survey responses were uploaded into SPSS version 22, and the responses to items 9 through 37 on the perceptual survey were reviewed. Responses to the Likert-type scales were assigned corresponding numbers of 1 through 5. The researcher used descriptive statistics in order to calculate the means and standard deviations of responses for the independent and dependent variables.

Qualitative Data Analysis

For Research Question 3, participants were given the opportunity to share their expertise with the researcher regarding the three variables of: (a) instructional practice scores, (b) value-added measures, and (c) learning gains. Item 38 asked the participants

in an open-ended format how they used the variables to make decisions in order to improve student scores for the 2013-2014 school year. The researcher categorized sets of responses and tallied the frequencies in order to apply the practical value of the three variables for the school leaders. Item 39 queried secondary principals and assistant principals for instruction as to their perceptions of what would improve the effectiveness of reading teachers. Again, the researcher categorized the responses in order to determine the frequency for further investigation (Lunenburg & Irby, 2008). The frequency of responses are discussed in detail in Chapter 4 and in relation to future research in Chapter 5.

Summary

The purpose of this chapter was to describe the variables principals must use in order to evaluate teacher effectiveness in the LUSD. The use of these metrics has been enacted into policy by both federal and state legislation beginning with the passing of the Race to the Top (U.S. Department of Education, 2009). Specifically, this study was conducted to analyze the relationship of the school based administrators' use of classroom observations, value-added scores assigned, and the learning gains obtained from the FCAT by secondary reading teachers (Florida State Statute 1012.34, 2012). The researcher further analyzed the extent to which school administrators utilized this information for decision-making as the instructional leaders of their schools.

In this chapter, a description of the target population was given. The roles of the participants were defined as either principals or assistant principals for instruction at

either the middle or high school levels. The data for Research Question 1 were derived from the iObservation and EDW databases within the LUSD (Large Urban School District, 2013b). Data for Research Questions 2 and 3 were obtained through a perceptual survey designed by the researcher. The instrument was validated by experts from the school district staff and faculty from the University of Central Florida. Procedures for the data collection methods for both quantitative and qualitative data were described, and response rates to the survey were included. Finally, statistical measures used in the analysis of data to respond to Research Questions 1, 2, and 3 were detailed.

CHAPTER 4 RESULTS

Introduction

This study focused on evaluating the relationships among ratings of teachers' instructional practices, value-added measures, and learning gains and the opinions of principals on the use or usefulness of these measures. Chapter 4 contains the results of the analysis of quantitative and qualitative data to answer the research questions which were used to guide the study. Results have been organized around the three research questions. Tables and accompanying narratives have been used to respond to each of the research questions.

Research Question 1

To what extent was there a relationship among each administrator's evaluation of teachers' instructional practices, the value-added measure, and the learning gains assigned to teachers from learning gains as measured by Florida Comprehensive Achievement Test (FCAT) Reading for the following grades: (a) 6 through 8 and (b) 9 through 12 in an urban school district for the school year 2012-2013?

H₀. There is no relationship between instructional practice scores attained through administrators' observation, the value-added measures, and the learning gains assigned to teachers based on the learning gains of their students in reading as determined by the FCAT in Grades 6 through 8 or 9 through 12.

H_A. There is a relationship among the instructional practice scores attained through administrators' observations, the value-added measures assigned to teachers, and the learning gains of their students in reading as determined by the FCAT in Grades 6 through 8 or 9 through 12.

The study school district provided a file with the instructional practice, value-added, and FCAT Reading learning gain scores for all teachers of students in Grades 6 through 10 for the 2012-13 school year. Only teachers with a valid score in all three categories were included in the study. For inclusion in the study, each reading teacher had to instruct enough students in a specified reading course to receive a value-added score calculation and learning gains calculation. When these criteria were placed on the original group of 955 reading teachers of students in Grades 6 through 10, a total of 883 teachers were available for the study.

Descriptive information on the different methods for measuring teacher effectiveness for the teachers included in the analysis is provided in Table 7. A total of 787 (89.1%) teachers in the analysis received an instructional practice score of Effective; 60 (6.8%) teachers received an instructional practice score of Highly Effective, and 36 (4.1%) teachers received an instructional practice score of Needs Improvement. No teachers in this group received an instructional practice rating of Unsatisfactory. A total of 537 (61%) teachers received a positive value-added score (greater than 0), indicating that they outperformed comparable teachers with comparable students. Of the teachers, 305 (34.5%) had a slightly negative score and 41 (4.5%) of teachers had a very negative score. The majority of teachers in the study (484) had between 60% and 75% of their

students make learning gains as calculated by the Florida Department of Education methods. These methods require proficient students to maintain their proficiency levels and for other students to meet a set gain in scale score points from prior to current year. A total of 98 (11.1%) reading teachers had less than 50% of their students making a learning gain.

Table 7

Descriptive Statistics for All Reading Teachers (N=883)

Descriptor	Frequency	Percentage
Instructional Practice Score		
Between 3.5 and 4.0 (Highly Effective)	60	6.8
Between 2.5 and 3.49 (Effective)	787	89.1
Between 1.5 and 2.49 (Needs Improvement)	36	4.1
Between 1.0 and 1.49 (Unsatisfactory)	0	0.0
Total	883	100.0
Value-added Score		
Between 0.5 and 3.0	70	8.0
Between 0.0 and 0.5	467	53.0
Between -0.5 and 0.0	305	34.5
Between -3.0 and -0.5	41	4.5
Total	883	100.0
Learning Gains		
Between 75% and 100%	83	9.4
Between 60% and 75%	484	54.9
Between 50% and 60%	218	24.6
Between 35% and 50%	95	10.7
Less than 35%	3	0.4
Total	883	100.0

Note. Percentage may not total 100% due to rounding.

A total of 25 middle school teachers (4.8%) were rated on instructional practice as Highly Effective with an additional 477 middle school teachers (91.4%) rated as Effective and 20 middle school teachers (3.8%) rated as Needs Improvement. Middle school teacher scores in instructional practice are more concentrated around the Effective rating than for all teachers. Similar to the instructional practice scores, the value-added scores for middle school teachers were more concentrated around '0' than all scores overall. A total of 23 middle school teachers (4.4%) had value-added scores between 0.5 and 3.0; 319 middle school teachers (61.1%) had value-added scores between 0.0 and 0.5; 171 middle school teachers (32.8%) had value-added scores between -0.5 and 0.0; and 9 middle school teachers (1.7%) had value-added scores between -0.5 and -3.0. Learning gains for middle school teachers were very similar to all teacher scores. A total of 41 middle school teachers (7.9%) had between 75% and 100% of their students make learning gains; 301 middle school teachers (57.7%) had between 60% and 65% of their students make learning gains; 126 middle school teachers (24.1%) had between 50% and 60% of their students make learning gains; 54 middle school teachers (10.3%) had between 35% and 50% of their students make learning gains; and no middle school teacher had fewer than 35% of their students make learning gains. Table 8 contains descriptive statistics for middle school and high school teachers (N=522).

Table 8

Descriptive Statistics for Middle School Reading Teachers (N=522)

Descriptor	Frequency	Percentage
Instructional Practice Score		
Between 3.5 and 4.0 (Highly Effective)	25	4.8
Between 2.5 and 3.49 (Effective)	477	91.4
Between 1.5 and 2.49 (Needs Improvement)	20	3.8
Between 1.0 and 1.49 (Unsatisfactory)	0	0.0
Total	522	100.0
Value-added Score		
Between 0.5 and 3.0	23	4.4
Between 0.0 and 0.5	319	61.1
Between -0.5 and 0.0	171	32.8
Between -3.0 and -0.5	9	1.7
Total	522	100.0
Learning Gains		
Between 75% and 100%	41	7.9
Between 60% and 75%	301	57.7
Between 50% and 60%	126	24.1
Between 35% and 50%	54	10.3
Less than 35%	0	0.0
Total	522	100.0

Note. Percentage may not total 100% due to rounding.

A total of 35 teachers (9.7%) were rated on instructional practice as Highly Effective. An additional 310 high school teachers (85.9%) were rated as Effective and 16 high school teachers (4.4%) were rated as Needs Improvement. High school teacher scores were more often Highly Effective and Needs Improvement than were the scores for all teachers. Similar to the instructional practice scores, the value-added scores for high school teachers were less concentrated around than the mean than for all scores overall. A total of 47 high school teachers (13.0%) had value-added scores between 0.5

and 3.0; 148 high school teachers (41.0%) had value-added scores between 0.0 and 0.5; 134 high school teachers (37.1%) had value-added scores between -0.5 and 0.0; and 32 high school teachers (8.9%) had value-added scores between -0.5 and -3.0. Learning gains for high school teachers were very similar to those of all teachers. A total of 42 high school teachers (11.6%) had between 75% and 100% of their students make learning gains; 183 school teachers (50.7%) had between 60% and 65% of their students make learning gains; 92 high school teachers (25.5%) had between 50% and 60% of their students make learning gains; 41 high school teachers (11.4%) had between 35% and 50% of their students make learning gains; and three high school teachers (0.8%) had fewer than 35% of their students make learning gains. Table 9 presents the descriptive statistics for high school teachers (N=361).

Table 9

Descriptive Statistics for High School Reading Teachers (N=361)

Descriptor	Frequency	Percentage
Instructional Practice Score		
Between 3.5 and 4.0 (Highly Effective)	35	9.7
Between 2.5 and 3.49 (Effective)	310	85.9
Between 1.5 and 2.49 (Needs Improvement)	16	4.4
Between 1.0 and 1.49 (Unsatisfactory)	0	0.0
Total	361	100.0
Value-added Score		
Between 0.5 and 3.0	47	13.0
Between 0.0 and 0.5	148	41.0
Between -0.5 and 0.0	134	37.1
Between -3.0 and -0.5	32	8.9
Total	361	100.0
Learning Gains		
Between 75% and 100%	42	11.6
Between 60% and 75%	183	50.7
Between 50% and 60%	92	25.5
Between 35% and 50%	41	11.4
Less than 35%	3	0.8
Total	361	100.0

Note. Percentage may not total 100% due to rounding.

To determine if a relationship existed between the three methods of teacher effectiveness, Pearson's r correlations were calculated for all combinations of the three methods. A standard critical value table was used to determine if statistical significance could be identified (Steinberg, 2010). Table 10 contains the results of the analysis.

Table 10

Pearson r Correlational Analysis Among Teacher Effectiveness Measures: All Teachers (N=883)

Effectiveness Measures	Instructional Practice Scores	Value-added Scores	Learning Gains
Instructional Practice Scores	-	0.01	-0.02
Value-added Scores	0.01	-	0.48**
Learning Gains	-0.02	0.48**	

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

There was no significant relationship between instructional practice (IP) scores and either of the quantitative assessment-based methods of measuring teacher effectiveness in the correlational analysis. The two quantitative measures, value-added scores and learning gains, were significantly related and the correlational relationship was moderate to strong (Steinberg 2010). This lack of a relationship indicated that though the learning gains and value-added measures were positively and significantly related to each other, neither measure had a meaningful correlation with the instructional practice scores provided by school administrators. A non-significant relationship with a sample size of 883 suggests no evidence of a relationship between these measures. This analysis partially failed to reject the null hypothesis that there is not a relationship between instructional practice scores, value-added scores, and learning gains.

There was a slightly stronger correlation between value-added scores and learning gains for middle school teachers as compared to high school teachers (.50 to .47), but this was not substantively different. The correlational analysis between the instructional practice scores and learning gains scores or value-added scores showed no substantive difference. Tables 11 and 12 present the correlational analyses for middle school and high school teachers.

Table 11

Correlational Analysis Among Teacher Effectiveness Measures: Middle School Teachers (N=522)

Effectiveness Measures	Instructional Practice Scores	Value-added Scores	Learning Gains
Instructional Practice Scores	-	0.01	-0.01
Value-added Scores	0.01	-	0.50**
Learning Gains	-0.01	0.50**	

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

Table 12

Correlational Analysis Among Teacher Effectiveness Measures: High School Teachers (N=361)

Effectiveness Measures	Instructional Practice Scores	Value-added Scores	Learning Gains
Instructional Practice Scores	-	0.01	-0.02
Value-added Scores	0.01	-	0.47**
Learning Gains	-0.02	0.47**	

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

Research Question 2

What factors do middle, and high school principals and assistant principals believe contribute to the relationships among the instructional practice ratings, the value-added measures, and the learning gains?

Principals provided information for their schools on percentage of students receiving free or reduced lunch (FRL) benefits, the percentage of students receiving special education services, and the percentage of students identified as English Language Learners.

Participating administrators' schools had sizable proportions of students participating in free and reduced lunch, special education, and English Language Learner programs. Over 70% of administrators worked at a school where 50% or more of students at their school participate in free and reduced lunch programs. Nearly 45% of administrators worked in schools where over 70% of students at their school participated in free and reduced lunch programs. Concerning students receiving special education services, over 49% of study schools had more than 16% of students receiving these services and over 26% of study schools had more than 21% of students receiving special education services. Over half of the schools had more than 11% of students participating in English Language Learner programs. Over 28% of schools had more than one-fifth of their students participating in English Language Learner programs. Table 13 displays student demographics of the participating principals' schools.

Table 13

Student Demographics: Participating Principals' Schools (N=116)

Descriptor	Frequency	Percentage
Students receiving free/reduced lunch benefits		
Less than 40%	16	13.8
41% to 50%	17	14.7
51% to 60%	12	10.3
61% to 70%	19	16.4
71% to 80%	10	8.6
81% to 100%	42	36.2
Total	116	100.0
Students receiving special education services		
Less than 5%	4	3.5
6% to 10%	21	18.1
11% to 15%	34	29.3
16% to 20%	26	22.4
21% to 25%	14	12.1
Above 25%	17	14.7
Total	116	100.0
Students identified as English language learners		
Less than 5%	21	18.1
6% to 10%	27	23.3
11% to 20%	35	30.2
21% to 30%	16	13.8
More than 30%	17	14.7
Total	116	100.0

Note. Percentage may not total 100% due to rounding.

Participating principals also shared personal and professional demographic information as to their gender, their highest degree earned, type of school, information on years of experience in their current position and total years of experience as an administrator. These data are displayed in Table 14.

Table 14

Personal and Professional Demographics: Participating Principals (N=116)

Descriptor	Frequency	Percentage
Gender		
Male	39	33.6
Female	77	66.4
Total	116	100.0
Highest degree earned		
Master's	77	66.4
Education Specialist	15	12.9
Doctorate	24	20.7
Total	116	100.0
Type of school		
Traditional	100	86.2
K-8	4	3.5
Exceptional Education Center	3	2.6
Alternative	8	6.9
Charter	1	.9
Total	116	100.1
Years of experience in current position		
Less than 3 years	58	50.0
3-5 years	23	19.8
6-10 years	27	23.3
11-15 years	6	5.2
More than 15 years	2	1.7
Total	116	100.0
Total years of experience as an administrator		
Less than 3 years	17	14.7
3-5 years	20	17.2
6-10 years	33	28.5
11-15 years	28	24.1
More than 15 years	18	15.5
Total	116	100.0

Note. Percentage may not total 100% due to rounding.

Participating administrators overwhelmingly had not spent a large portion of their careers in their current position. Of the 116 administrators responding to the survey, 58 (50%) had less than three years of experience in their current positions, and only 8 (6.9%) had over a decade of experience in their current positions. Overall, however, participants were very experienced administrators. A total of 79 (68.1%) administrators had more than six years of administrative experience in their careers and 46 (39.6%) administrators had more than a decade of administrative experience. All 116 participants in the study had at least a Master's degree with 15 (12.9%) holding an additional Educational Specialist degree and 24 (20.7%) having earned a doctoral degree. Females (77, 66.4%) outnumbered males (39, 33.6%) two to one in the study. Almost twice as many participants were female as male and about one-third had an advanced graduate degree beyond the master's degree. Nearly all participating administrators (100, 86.2%) worked in traditional schools.

For the alignment between instructional practice scores and value-added scores, a plurality of administrators (47, 44.3%) agreed that instructional practice scores and value-added measures were aligned. However, no administrators strongly agreed that there was alignment between the two measures, and a total of 45 (42.4%) of the administrators disagreed or strongly disagreed that there was alignment between the two measures. The remaining 14 (13.2%) indicated that they did not know if alignment existed. Administrators, therefore, were nearly equally divided on the overall question of whether instructional practice scores and value-added measures were aligned. Table 15 displays administrators' opinions on the use of instructional practice scores in summative

evaluations and the perceptions of administrators on their usefulness compared to learning gains and value-added scores.

Table 15

Administrators' Opinions: Usefulness of Instructional Practice Scores (N=106)

Survey Stem (N)	Strongly Agree <i>f</i> (%)	Agree <i>f</i> (%)	Disagree <i>f</i> (%)	Strongly Disagree <i>f</i> (%)	Do Not Know <i>f</i> (%)
Instructional Practice scores...					
are aligned to value-added measures. (106)	0 (0.0)	47 (44.3)	42 (39.6)	3 (2.8)	14 (13.2)
are aligned to learning gains of students. (106)	0 (0.0)	58 (54.7)	37 (34.9)	4 (3.8)	7 (6.6)
should be used in summative evaluation. (106)	14 (13.2)	70 (66.0)	13 (12.3)	1 (.9)	8 (7.6)
are more reflective of overall performance than learning gains. (106)	4 (3.8)	35 (33.0)	52 (49.1)	1 (.9)	14 (13.2)
are more reflective of overall performance than value-added scores. (106)	12 (11.3)	47 (44.3)	33 (31.1)	1 (.9)	13 (12.3)

A small majority of administrators (58, 54.7%) expressed the belief that instructional practice scores were aligned with the learning gains of students. No administrators, however, stated that they strongly agreed with this statement. This finding indicated that there may not be high levels of certainty about the alignment between learning gains and instructional practice scores. Despite the majority agreeing that there was alignment, 41 (38.6%) of administrators either disagreed or strongly disagreed that there was alignment between instructional practice scores and the learning gains of students.

Administrators overwhelmingly agreed that instructional practice scores should be used in the summative evaluation. Of the respondents, 84 (79.3%) either agreed or strongly agreed that instructional practice scores belong in final evaluation ratings. In addition, only one administrator, less than 1%, strongly disagreed that instructional practice scores should be used.

Despite agreeing that instructional practice scores should be used, administrators did not think that instructional practices scores were more reflective of a teacher's overall performance than learning gains. A total of 53, exactly 50% of responding administrators, disagreed or strongly disagreed that instructional practice scores were more reflective of a teacher's overall performance than learning gains. Only 39 (36.8%) believed that instructional practice scores were more reflective than learning gains.

In regard to value-added scores, 59 (55.6%) responding administrators strongly agreed or agreed that instructional practice scores were more reflective of overall

performance than value-added scores. A substantial number of responding administrators (34, 32%) disagreed or strongly disagreed, and 13 (12.3%) indicated they were unsure.

A plurality of administrators (46, 46.9%) believed that value-added scores were aligned with the learning gains of students. A total of 40 administrators (40.9%) either disagreed or strongly disagreed that value-added scores were aligned with the learning gains of students. An additional 12 administrators (12.2%) indicated that they did not know if alignment existed between the value-added scores and learning gains.

Administrators provided similar feedback on the alignment of value-added scores with instructional practice scores. A small plurality of 44 administrators (44.9%) agreed that there was alignment. In contrast, 40 administrators (40.9%) disagreed or strongly disagreed that there was alignment. A total of 14 (14.3%) administrators answered that they did not know if there was alignment between value-added and instructional practice scores.

Even with disagreement over the alignment of the scores, a total of 54 administrators (55.1%) stated that value-added scores should be used in the summative evaluation. Opposing the use of value-added scores in the summative evaluation were 34 administrators (34.7%) who disagreed or strongly disagreed that value-added scores should be used. Despite supporting the use of value-added scores in the summative evaluation, a majority of administrators stated that the value-added scores were not more reflective of teachers' overall performance than either instructional practice scores or learning gains. A total of 61 administrators (62.2%) disagreed or strongly disagreed that value-added scores were more reflective of overall teacher performance than instructional

practice scores, and 62 administrators (63.2%) disagreed or strongly disagreed that value-added scores were more reflective of overall teacher performance than learning gains. A total of 13 administrators (13.3%) and 11 administrators (11.2%) did not know whether value-added scores were more or less reflective of overall teacher performance than instructional practice scores. Table 16 displays administrators' opinions on the use of value-added scores.

Table 16

Administrators' Opinions: Usefulness of Value-added Scores (N=98)

Survey Stem (N)	Strongly Agree <i>f</i> (%)	Agree <i>f</i> (%)	Disagree <i>f</i> (%)	Strongly Disagree <i>f</i> (%)	Do Not Know <i>f</i> (%)
Value-added scores... are aligned to the learning gains of students. (98)	1 (1.0)	45 (45.9)	37 (37.8)	3 (3.1)	12 (12.2)
are aligned to instructional practice scores. (98)	0 (0.0)	44 (44.9)	37 (37.8)	3 (3.1)	14 (14.3)
should be used in summative evaluation. (98)	3 (3.1)	51 (52.0)	29 (29.6)	5 (5.1)	10 (10.2)
are more reflective of overall performance than instructional practice. (98)	2 (2.0)	22 (22.5)	51 (52.0)	10 (10.2)	13 (13.3)
are more reflective of overall performance than learning gains. (98)	0 (0.0)	25 (25.5)	55 (56.1)	7 (7.1)	11 (11.2)

Because questions concerning the relationships between learning gains and other measures were asked in earlier sections, they were not repeated in this section. A majority of administrators (57, 58.8%) stated that the learning gains of students reflected instructional practice scores. A total of 36 administrators (37.1%) disagreed, and an additional four administrators (4.1%) stated that they did not know.

There was strong consensus among administrators concerning the use of learning gains in the summative evaluation. A large majority (74, 76.3%) strongly agreed or agreed that learning gains should be used as a portion of the summative evaluation. Only 16 administrators (16.5%) believed that learning gains should not be used in the summative evaluation. Learning gains emerged as the most strongly supported element among the three methods of evaluating teacher effectiveness described in this survey. Table 17 displays administrators' opinions as to the use of learning gains.

Table 17

Administrators' Opinions: Usefulness of Learning Gains Scores (N=97)

Survey Stem (N)	Strongly Agree <i>f</i> (%)	Agree <i>f</i> (%)	Disagree <i>f</i> (%)	Strongly Disagree <i>f</i> (%)	Do Not Know <i>f</i> (%)
Learning Gains scores...					
reflect instructional practice scores. (97)	0 (0.0)	57 (58.8)	36 (37.1)	0 (0.0)	4 (4.1)
should be used in summative evaluation. (97)	10 (10.3)	64 (66.0)	15 (15.5)	1 (1.0)	7 (7.2)

Overall, administrators expressed highest confidence in the use of learning gains for summative evaluations as compared to instructional practice scores or value-added scores. Administrators were least confident in the use of value-added scores. Despite different levels of confidence, a majority of administrators supported the use of all three measures in a teacher's summative evaluation.

Research Question 3

To what extent do principals report using the instructional practices score, learning gains, or VAM scores to make personnel or instructional decisions?

Instructional Practice Scores

A total of 59 administrators (55.6%) agreed or strongly agreed that instructional practice scores were used in making teacher assignments. A minority of administrators (34, 32.0%) disagreed that instructional practice scores were used in making teacher assignments. An additional 13 administrators (12.3%) did not know if instructional practice scores were used.

A large majority of administrators (77, 72.7%) agreed or strongly agreed that instructional practice scores help target professional learning needs. Less than a quarter of administrators (22, 20.8%) disagreed that instructional practice scores helped to target professional learning needs. No administrators strongly disagreed that instructional practice scores helped to target professional learning needs, and seven administrators (6.6%) did not know if these scores helped.

A sizeable minority of administrators (43, 40.5%) agreed or strongly agreed that instructional practice scores differed based on the administrator conducting the evaluation. A total of 50 administrators (47.2%) disagreed that instructional practice scores differed based on the administrator conducting the evaluation. An additional 13 administrators (12.3%) did not know if the scores differed by administrator. The information in Table 18 displays administrators' responses about the use of instructional practice scores to make decisions.

Table 18

Administrators' Opinions: Use of Instructional Practice Scores to Make Decisions (N=106)

Survey Stem (N)	Strongly Agree <i>f</i> (%)	Agree <i>f</i> (%)	Disagree <i>f</i> (%)	Strongly Disagree <i>f</i> (%)	Do Not Know <i>f</i> (%)
Instructional Practice scores...					
are used in making teaching assignments. (106)	12 (11.3)	47 (44.3)	33 (31.1)	1 (.9)	13 (12.3)
vary depending on the administrator conducting the evaluation. (106)	10 (9.4)	33 (31.1)	50 (47.2)	0 (0.0)	13 (12.3)
help target professional learning needs. (106)	15 (14.2)	62 (58.5)	22 (20.8)	0 (0.0)	7 (6.6)
give valuable data for meaningful discussions with teachers. (106)	18 (17.0)	58 (54.7)	24 (22.6)	0 (0.0)	6 (5.7)

A total of 103 administrators (97.2%) either agreed or strongly agreed that they were confident in their ability to conduct instructional practice evaluations. One administrator strongly disagreed, and two administrators expressed uncertainty regarding their level of confidence in their ability to conduct evaluations. Table 19 reveals that administrators overall had strong confidence in their ability to conduct evaluations but

less confidence in the ability of other administrators at their schools to conduct evaluations that were consistent with those of other administrators.

Table 19

Administrators' Confidence in the Use of Instructional Practice Scores (N=106)

Survey Stem (N)	Strongly Agree <i>f</i> (%)	Agree <i>f</i> (%)	Disagree <i>f</i> (%)	Strongly Disagree <i>f</i> (%)	Do Not Know <i>f</i> (%)
I am confident in my ability to conduct instructional practice evaluations. (106)	43 (40.6)	60 (56.6)	0 (0.0)	1 (.9)	2 (1.9)

Open-ended Responses: Instructional Practice Scores

A total of 13 administrators provided open-ended responses concerning the use of instructional practice scores to improve reading instruction. Six administrators (46.2%) described the instructional practice scores positively. A total of four administrators (30.8%) had a mixed response to the instructional practice scores and three administrators (23.0%) had a negative response to the instructional practice scores. These data are displayed in Table 20.

Also displayed in Table 20 is information as to the frequency with which administrators who mentioned instructional practice scores described specific use of the scores in school-based decisions. Administrators could state that they used instructional

practice scores to make more than one category of decisions. Six of the 13 administrators (46.2%) stated that they used instructional practice scores in making decisions about professional learning. A total of four administrators (30.8%) described their use of instructional practice scores in making decisions about teacher assignment. Only one administrator (7.7%) described using the instructional practice scores to make decisions concerning material selection.

Table 20

Administrators' Open-ended Responses: Instructional Practice Scores (N=13)

Descriptor	Frequency	Percentage
Use of evaluation data to improve reading instruction		
Positive Response	6	46.2
Mixed Response	4	30.8
Negative Response	3	23.0
Total	13	100.0
Use of evaluation element for selected school-based decisions		
Professional development	6	46.2
Teacher assignment	4	30.8
Material selection	1	7.7
No specific decision provided	2	15.4
Total	13	100.0

Administrators' open-ended responses expanded on their survey responses. One administrator stated that "Though instructional practice scores are inflated in general, they are at least immediate and identify the strengths and weaknesses of the teacher." This comment described both the usefulness of the instructional practice scores for school decision-making along with concerns about potential inflation and inconsistency in scores

throughout the school. Another administrator believed that the use of instructional practice will continue strengthen as additional elements are added. This administrator stated, “Now that all elements are available as a part of the instructional practice score, the review of data in the future will more heavily include their instructional practice data.” Another administrator echoed this opinion, adding that increased experience with the instructional practice framework over time would also help increase the confidence that other administrators are coming to common decisions about evaluation ratings. This administrator commented that “The administrators were very inexperienced with all 41 elements in the Marzano program for this year, and we expect improvement next year.” At the time of the study, the large, urban school district had only one year of experience in the use of all 41 elements in the Domain 1 portion of the Marzano evaluation system, having adopted that system for the first time at the beginning of the 2013-14 school year.

Value-added Scores

A large majority of administrators (60, 61.2%) disagreed or strongly disagreed that value-added scores were used in making teacher assignments at their school. Only 30 administrators (30.6%) agreed or strongly agreed that these scores were used to inform teacher assignments. An additional eight administrators (8.2%) did not know how value-added scores were used to inform teacher assignments.

A total of 30 administrators (30.6%) expressed that value-added scores assigned to teachers aligned as they had expected. A total of 52 administrators (53.2%) disagreed that the value-added scores were as they expected, and an additional 16 administrators

(16.3%) did not know. Additionally, 42 administrators (42.9%) agreed or strongly agreed that the value-added scores assigned to teachers were fair. A majority of administrators (50, 51.0%) disagreed or strongly disagreed that the value-added scores awarded to teacher were fair. Six administrators (6.1%) did not know if the scores awarded were fair.

Concerning the use of value-added data for making school decisions, only 25 administrators (25.5%) stated that value-added data helped to target professional learning needs. A large majority of administrators (62, 63.2%) disagreed or strongly disagreed that value-added scores helped with professional learning needs, and 11 administrators (11.2%) stated that they did not know. A small majority of administrators (50, 51.0%) stated that value-added scores did give valuable information for meaningful discussions with teachers. A total of 32 administrators (32.6%) expressed that value-added scores did not give valuable information for meaningful discussions, and 16 administrators (16.3%) responded that they did not know if the scores gave valuable information. Table 21 displays administrators' responses on the use of value-added scores to make decisions.

Table 21

Administrators' Opinions: Use of Value-added Scores to Make Decisions (N=98)

Survey Stem (N)	Strongly Agree <i>f</i> (%)	Agree <i>f</i> (%)	Disagree <i>f</i> (%)	Strongly Disagree <i>f</i> (%)	Do Not Know <i>f</i> (%)
Value-added scores...					
are used in making teaching assignments. (98)	2 (2.0)	28 (28.6)	49 (50.0)	11 (11.2)	8 (8.2)
assigned to teachers were what was expected. (98)	0 (0.0)	30 (30.6)	50 (51.2)	2 (2.0)	16 (16.3)
help target professional learning needs. (98)	0 (0.0)	25 (25.5)	55 (56.1)	7 (7.1)	11 (11.2)
give valuable data for meaningful discussions with teachers. (98)	0 (0.0)	50 (51.0)	30 (30.6)	2 (2.0)	16 (16.3)
assigned to teachers were fair. (98)	3 (3.1)	39 (39.8)	43 (43.9)	7 (7.1)	6 (6.1)

Open-ended Responses for Value-Added Scores

Twenty administrators provided open-ended responses concerning value-added scores. A single administrator (5.0%) described having a positive response to using the value-added scores to improve reading instruction. A total of three administrators (15%) provided a mixed response, while the remaining 16 administrators (80%) provided a

negative response. Of these 20 administrators, only 2 administrators (10%) reported using the value-added scores in school-based decisions, specifically for informing professional learning. Table 22 displays the open-ended responses of administrators regarding the use of evaluation data to improve reading instruction and the use of evaluation elements for selected school-based decisions.

Table 22

Administrators' Open-ended Responses: Value-added Scores (N=20)

Descriptor	Frequency	Percentage
Use of evaluation data to improve reading instruction		
Positive Response	1	5.0
Mixed Response	3	15.0
Negative Response	16	80.0
Total	20	100.0
Use of evaluation element for selected school-based decisions		
Professional development	2	10.0
Teacher assignment	0	0.0
Material selection	1	5.0
No specific decision provided	17	85.0
Total	20	100.0

The open-ended responses by administrators provided supporting evidence for the quantitative survey responses. Administrators described the timing of value-added scores as a strong concern in their usefulness. One administrator stated, “Value-added measures come out so late that we have already made staffing decisions by the time that they are known.” Another administrator observed that “The value-added scores were meaningless for planning for the next school year since they did not come out until we were 4 months

into the year.” Administrators consistently described (a) learning gains information available at the conclusion of the statewide assessment and (b) instructional practice evaluations concluded by the end of the school year as integral in their summative determinations.

Administrators also expressed either frustration or methodological concerns in regard to the value-added measure. Some administrators suggested that since they were unable to understand and explain the value-added method, they were unable to use it. One administrator described the value-added model as “a complex calculation that is difficult for teachers to fully understand.” Other administrators were more forthright in their concerns, describing value-added as a “joke” or “unreliable.” Administrators also described the difficulty in describing for teachers value-added measures based on school-level scores rather than the scores of their specific students.

Learning Gains

Over three-quarters of administrators (76, 78.4%) agreed or strongly agreed that learning gains reflected the quality of reading instruction. A total of 18 administrators (18.6%) disagreed that learning gains were reflective of quality instruction, and an additional three administrators (3.1%) indicated they did not know.

A total of 70 administrators (72.2%) agreed or strongly agreed that learning gains were used in making teaching assignments. Only 23 administrators (23.7%) disagreed or strongly disagreed, and four administrators (4.1%) responded that they did not know if learning gains were used for making teaching assignments. Nearly all administrators (81,

83.5%) agreed or strongly agreed that learning gains were used to help target professional learning needs. A total of 12 administrators (12.4%) disagreed that learning gains were used to help target professional learning needs, and four administrators (4.1%) reported that they did not know.

A total of 84 administrators (86.6%) agreed or strongly agreed that learning gains gave valuable data for meaningful conversations with teachers. Fewer than 1 in 10 administrators (9, 9.3%) believed that learning gains were not valuable for meaningful conversations and four administrators (4.1%) responded that they did not know.

Administrators also reported that learning gains helped to prompt reviews of scope and sequence of curriculum along with reviews of instructional materials. A total of 59 administrators (60.9%) stated that learning gains prompted a review of the scope and sequence. One-third of responding administrators (32, 33.0%) agreed or strongly agreed that learning gains prompted a review, and six administrators (6.2%) indicated that they did not know.

A total of 79 administrators (81.5%) agreed or strongly agreed that learning gains prompted a review of how time was utilized in the classroom. An additional 13 administrators (13.4%) disagreed that a review of time was prompted, and five administrators (5.2%) reported that they did not know. Table 23 displays administrator opinions about the use of learning gains to make decisions.

Table 23

Administrators' Opinions: Use of Learning Gains to Make Decisions (N=97)

Survey Stem (N)	Strongly Agree <i>f</i> (%)	Agree <i>f</i> (%)	Disagree <i>f</i> (%)	Strongly Disagree <i>f</i> (%)	Do Not Know <i>f</i> (%)
Learning gains scores...					
reflect the quality of the reading instruction. (97)	5 (5.2)	71 (73.2)	18 (18.6)	0 (0.0)	3 (3.1)
were used in making teaching assignments. (97)	6 (6.2)	64 (66.0)	22 (22.7)	1 (1.0)	4 (4.1)
help target professional learning needs. (97)	14 (14.4)	67 (69.1)	12 (12.4)	0 (0.0)	4 (4.1)
give valuable data for meaningful discussions with teachers. (97)	18 (18.6)	66 (68.0)	9 (9.3)	0 (0.0)	4 (4.1)
prompted you to review the scope and sequence in classes. (97)	8 (8.3)	51 (52.6)	31 (32.0)	1 (1.0)	6 (6.2)
prompted you to review instructional materials. (97)	15 (15.5)	57 (58.8)	19 (19.6)	1 (1.0)	5 (5.1)
prompted you to review how time was utilized. (97)	21 (21.7)	58 (59.8)	12 (12.4)	1 (1.0)	5 (5.2)

Open-ended Responses for Learning Gains

A total of 27 administrators provided open-ended responses concerning learning gains scores. Only one administrator (3.7%) described learning gains scores with a mixed response. All other administrators (26, 96.3%) described learning scores positively in their use for improving reading instruction. Seven (25.9%) of the administrators providing open-ended responses concerning learning gains scores reported that they used these scores to inform professional learning. Twelve administrators (44.4%) described using learning gains in the assignment of teachers, including the assignment of teachers to particular subjects and grade levels. An additional two administrators (7.4%) used learning gains in the selection of instructional materials.

Learning gains were praised nearly universally by administrators in their open responses. This praise was often in contrast to information received from value-added scores. One administrator stated that in comparison to value-added scores, “Learning gains are more valid in assessing a teacher’s strength and weaknesses.” Directly addressing the idea of timing and ease of use for decision-making, another administrator stated, “Learning gains are easier to understand and calculate and are made public early in the summer to be used for staffing decisions.” Three administrators used “valid” to describe learning gains and “invalid” in reference to value-added scores or other measures. Administrators also valued the student-level data that were provided by learning gains as opposed to the aggregate school and teacher ratings resulting from value-added data. The student-level data were viewed as more accessible to teachers and, therefore, potentially useful in facilitating further discussion.

Table 24 displays the open-ended responses of administrators regarding the use of evaluation data to improve reading instruction and the use of evaluation elements for selected school-based decisions.

Table 24

Administrators' Open-ended Responses: Learning Gains Scores (N=27)

Descriptor	Frequency	Percentage
Use of evaluation data to improve reading instruction		
Positive Response	26	96.3
Mixed Response	1	3.7
Negative Response	0	0.0
Total	27	100.0
Use of evaluation element for selected school-based decisions		
Professional development	7	25.9
Teacher assignment	12	44.4
Material selection	2	7.4
No specific decision provided	6	22.2
Total	27	100.0

Note. Percentages may not total 100% due to rounding.

*Administrators' Open-Ended Responses:
Strategies to Improve the Effectiveness of Reading Teachers*

A total of 52 administrators provided additional open-ended responses, sharing their opinions concerning strategies that would improve the effectiveness of reading teachers. Table 25 contains a summary of the analysis of the administrators' responses.

Table 25

Administrators' Open-ended Responses: Strategies to Improve Reading Teachers' Effectiveness (N=52)

Strategies	Frequency	Percentage
Increased professional learning	11	21.2
Increased classroom flexibility for teachers	8	15.4
Recruiting better reading teachers	6	11.5
Training on instructional software program	5	9.6
Smaller class sizes	4	7.7
Use instructional practice evaluation for coaching instead of evaluation	3	5.8
Student motivation	2	3.8
Increased rigor in instructional practice evaluation	2	3.8
Increased training for assisting ESE students	1	1.9
Focused core subjects	1	1.9
Change deliberate practice calculation	1	1.9
Remove ineffective reading teachers	1	1.9
Strengthen summer reading program	1	1.9
Improve reading endorsement program	1	1.9
District-level meetings of reading teachers	1	1.9
Video-taping lessons	1	1.9
Increased use of informational text	1	1.9
More money for materials	1	1.9
Avoiding shifting reading programs	1	1.9
Total responses	52	99.7

Note. Percentage may not total 100% due to rounding.

Though no specific recommendation was provided by a majority or near-majority of administrators, the largest number of administrators (11, 21.2%) stated that increased professional learning would help to improvement the effectiveness of reading teachers. This suggestion was often provided in a list of suggestions with no additional information. One detailed suggestion in this category involved a perceived need for professional learning on how to incorporate more rigorous lessons with intensive reading

students. Three of the administrators mentioned that this type of professional learning opportunities was needed by both reading teachers and reading coaches.

Eight administrators (15.4%) identified the need to allow increased classroom flexibility for teachers. This was most often expressed in terms of allowing exceptions to reading instructional models that relied on the use of an instructional software program. One administrator described allowing for “teachers with proven results to deviate from prescriptive programs without micromanagement from a company.” There were concerns in other comments about a potential loss of teacher autonomy. Loss of autonomy was linked to making reading instruction less appealing to teachers.

A total of six administrators (11.5%) expressed a need for recruiting more and better reading teachers. Some administrators believed that paying bonuses for performance or staying for a period of time in reading would help. For example, one administrator stated that “A monetary incentive may help bring good teachers to teach reading courses.” Another administrator commented on the difficulty of the job and the need for a broader pool of secondary level reading teachers. Respondents stated that “Reading teachers have a difficult job and often even the best teachers can get burnt out after a few years with the demands placed on them.”

Between 5% and 10% of administrators stated that professional learning on instructional software programs (9.6%), smaller class sizes (7.7%), and changing the focus of the instructional practice system to coaching from evaluation (5.8%) would be helpful. Two administrators (3.8%) proposed that student motivation and increasing the rigor of the instructional practice evaluation would improve the effectiveness of reading

teachers. Other responses included increased instructional strategy training in teaching ESE students, increased focus on reading strategies for teachers in core subjects, changing the deliberate practice calculation, removing ineffective reading teachers, strengthening summer reading programs, improving the reading endorsement program, holding school district-level meetings of reading teachers, taping lessons, increasing the use of informational text, allocating more money for reading materials, and avoiding school district changes in reading programs.

Summary

The analyses of the data have been presented in this chapter. Quantitative and qualitative data were used to respond to the three research questions that guided the study. Following are brief summaries of the results of the analyses conducted for each of the questions.

In response to Research Question 1, no statistically significant correlation was found between the instructional practice measure and either of the quantitative, assessment-based measures of measuring teacher quality. A moderate to strong correlation existed between the two quantitative measures of teacher quality: learning gains and value-added scores.

For Research Question 2, administrators were most confident in the use of learning gains scores and they believed these to be most appropriate for use in the evaluation of teachers. Second in level of confidence expressed by administrators was the measure of instructional practice scores. Administrators were least confident in the

use of value-added scores. However, administrators believed that there is a place for all three methods in constructing summative evaluations for teachers.

For Research Question 3, administrators indicated that they preferred the use of learning gains to make decisions about teacher placement and professional learning opportunities along with other school decision-making processes. Both instructional practices and learning gains data were available in time to make decisions, but value-added data were released at the beginning of the following school year, which is much too late to inform staffing and other decisions for the upcoming school year. Administrators also expressed other concerns regarding value-added data that directly impacted their trust and confidence in the use of these data.

CHAPTER 5 SUMMARY, DISCUSSION, AND IMPLICATIONS

Introduction

This study was initiated to analyze the impact of the current evaluation system used in the Large Unit School District (LUSD) for secondary reading teachers by school administrators that was implemented in compliance with Florida State Statute 1012.34. This final chapter of the dissertation has been organized into the following four sections. The first section contains a summary of the study. The second section provides a discussion of the results of the quantitative and qualitative data analyses organized around the three research questions which guided the study. Implications for practice for the LUSD and other school districts facing similar legislation involving value-added measures, instructional practice scores, and learning gains are presented in the third section. The chapter is concluded with recommendations for further research and a final summary statement.

Summary of the Study

As a result of Race to the Top and Florida State Statute 1012.34 (2012), school districts within the state of Florida were required to adopt an evaluation system for teachers and school administrators based on student growth and teacher observations (U.S. Department of Education, 2012). Under 1012.34 (2012), Florida school districts had to adopt an evaluation system that linked compensation for teachers and principals to their annual evaluation. Likewise, tenure was eliminated for teachers not awarded it prior

to July 1, 2011. Finally, the value added by each teacher had to be determined by 2014-2015 for every course taught in Florida schools (Florida State Statute 1012.34, 2012). Consequently, the LUSD, in which the study was conducted, adopted the Marzano evaluation system.

The purpose of the study was to determine the relationship that existed among instructional practice scores, value-added scores, and learning gains earned by secondary reading teachers. For the LUSD, the intent of the study was to provide feedback regarding areas requiring further focus for school administrators regarding the evaluation process for secondary reading teachers.

Three research questions were addressed during this study. They focused on (a) the extent to which there was a relationship among each administrator's evaluation of teachers' instructional practices, the value-added measure, and the learning gains assigned to teachers from learning gains as measured by Florida Comprehensive Achievement Test (FCAT) Reading for Grades 6-8 and 9-12 in an urban school district for the school year 2012-2013; (b) the factors surveyed administrators believed contributed to the relationships among the instructional practice ratings, the value-added measures, and the learning gains; and (c) the extent to which principals reported using the instructional practices score, learning gains, or VAM scores to make personnel or instructional decisions.

The administrator population for the study consisted of 275 principals and assistant principals for instruction at the middle and high school levels for LUSD for the 2012-2013 school year. Specifically, the population consisted of 81 secondary principals

and 194 assistant principals. The purposeful sample consisted of 138 school based administrators, 65 principals and 73 assistant principals for instruction, who completed classroom observations on secondary reading teachers. Data to respond to Research Question 1 were secured from the iObservation instrument and the school district's educational warehouse. From the analysis, the significance of the relationships were determined that existed among instructional practice scores, value-added measures, and learning gains. To gather data to respond to Research Questions 2 and 3, all 138 participants were requested to complete the Instructional Practice, Value-Added Measure, Secondary Learning Gains Survey (Appendix D). After four contacts with potential respondents, the final usable return rate for the 116 responding administrators was 84%.

Research Question 1 was addressed by calculating the linear relationship that existed between the variables of instructional practice scores, value-added measures, and learning gains. For the analysis, the teachers had to have valid scores for all three areas. Consequently, 883 teachers were used in the analysis.

Research Question 2 focused on the perceptions of the participants regarding instructional practice scores, value-added measures, and learning gains. The survey instrument (Appendix D) contained specific sections for all three variables. Participant beliefs were calculated for each survey item using percentages and a five-point Likert type scale.

Research Question 3 targeted the extent to which school administrators used instructional practice scores, value-added measures, and learning gains to make personnel and instructional decisions. Percentages were calculated for each item regarding

participant agreement with the scripted survey statements. The intent of the final two survey items was to solicit open-ended responses from the participants regarding pertinent information that might not have been addressed by previous items.

Consequently, participants were able to state how they personally used the 2012-2013 data from instructional practice scores, value-added measures, and learning gains to improve reading performance at their schools. In addition, participants were afforded the opportunity to express their overall suggestions for improving the effectiveness of reading teachers.

Discussion of the Findings

Research Question 1

To what extent was there a relationship among each administrator's evaluation of teachers' instructional practices, the value-added measure, and the learning gains assigned to teachers from learning gains as measured by Florida Comprehensive Achievement Test (FCAT) Reading for the following grades: (a) 6 through 8 and (b) 9 through 12 in an urban school district for the school year 2012-2013?

A correlation coefficient was calculated for each possible linear relationship of the following variables: (a) instructional practice scores, (b) value-added measures, and (c) learning gains. Results indicated that over 95% of instructional practice scores for the secondary reading teachers were either "Effective" or "Highly Effective" on the summative instrument. These evaluation classifications were assigned based on the

classroom observations ratings assigned by assistant principals or principals of the schools. This result was aligned with previous studies such as the Widget Effect Study conducted in Chicago schools where 99.5% of all teachers were rated at the Satisfactory level or above as determined by classroom observations from school administrators (Bill & Melinda Gates Foundation, 2010). For the LUSD, less than 4% of the secondary reading teachers earned a “Needs Improvement” on the instructional practice portion of their evaluations. Furthermore, no secondary reading teachers earned an “Unsatisfactory” as determined by administrative observations.

These results indicate that a need exists for the LUSD to invest in professional learning for school administrators, focusing on conducting effective classroom observations. According to Marshall (2012), sufficient training for school administrators accompanied by a certification of proficiency, establishes a standard of competency for administrators prior to actually conducting classroom observations. These data also suggest that the administrators of the LUSD may require assistance in having courageous conversations with their teachers. School leaders must be skilled communicators, thereby ensuring that the process of improving instruction is an interactive venue focused on accurate reflection of what is observed (Marzano, 2007).

As revealed in the data analysis, there was a discrepancy in the percentage of teachers receiving positive value-added scores and those earning “effective” or “highly effective” ratings through administrative observations. The value-added measure is designed to allow school districts to quantify the added value of teachers’ instruction to their specific students’ outcomes (Sanders & Horn, 1998). For the LUSD, 61% of the

secondary reading teachers earned a positive value-added score in the area of reading, indicating that they outperformed comparable teachers with comparable students. The majority of the teachers with positive value-added scores earned between 0 and .5. Approximately 8% of the teachers had significant growth, ranging from .5 to 3.0. Conversely, 35% of the secondary reading teachers earned negative value-added scores as compared to only 4% who earned a “needs improvement” on the instructional practice portion of the evaluation. The actual Pearson r displayed a .01 correlation coefficient for instructional practice scores and value-added measures. This clearly indicated that there was not a significant relationship between the instructional practice scores assigned by school administrators and the value-added measures earned by secondary reading teachers. Administrators did not express confidence in the validity of the value-added scores making it unlikely that the lack of alignment between these measures would drive administrators to use value-added scores to calibrate their instructional practice scores.

Additional challenges can be identified for the value-added scores assigned to secondary reading teachers in the LUSD. First, value added models have been used to display student growth when students are randomly assigned to teachers (Newton et al., 2010). For students being served in secondary reading courses, their placement is often purposeful because of low academic performance for the subsequent year. For the secondary reading teachers, every student in their courses performed below the proficiency level, as measured on the FCAT Reading, for the preceding school year. Furthermore, depending on students’ levels of performance on the FCAT, they may receive a single or double block of reading. This variation in amount of instructional

time and the structure for delivering the curriculum can cause a considerable challenge for secondary schools and the use of value-added models (Newton et al., 2010). Value-added models such as the one used in Florida attempt to control for this through the use of covariates for the number of subject-specific courses taken by students in the academic year (Florida Department of Education, 2012). Even with these adjustments, however, the model still is unable to control for large differences in scheduling practices, supplemental programs, and the extension of the school day through tutoring and other student supports.

When analyzing the relationship of instructional practice scores and learning gains, the data reveals similar findings. There is a significant discrepancy between instructional practice scores and learning gains. The Pearson r showed a linear relationship of $-.02$ between the two variables. A majority of the secondary reading teachers had between 60% and 75% of their students make learning gains as measured by FCAT. Approximately 35% of the teachers had 50% or less of their students earn learning gains for the 2012-2013 school year. Given the confidence of administrators in learning gains scores, administrators may be amenable to using these scores to strengthen the rigor of the instructional practice scores. According to Chin et al. (2010), the FCAT items have been scaled and calibrated appropriately for the accompanying grade levels. This same validity was not supported by administrative observation studies. According to Fink (2012), expertise with pedagogy is the best predictor of administrative accuracy regarding classroom observations. Furthermore, Weisburg et al. (2009) noted that

observations from school leaders must reflect accurate daily instruction in order to be meaningful.

The results of the analysis of this portion of Research Question 1 indicate that the LUSD should continue to invest in building the skills of school administrators regarding their ability to conduct accurate classroom observations. Emphasis should be placed on inter-rater reliability training for all current administrators and on embedding such training in preparation programs for aspiring leaders.

The strongest relationship among the three variables was between value-added measures and learning gains with a Pearson r of .48. The covariate adjustment model incorporates the use of prior FCAT scores into its calculation (Florida Department of Education, 2014c). Therefore, it is logical that a strong relationship exists among these variables. This relationship exists in spite of differences between the two models that may make the calculations particularly inconsistent for some groups of teachers. Breaking up the analysis by middle school and high school teachers did not find significant differences between the correlational analysis of these two groups.

The learning gains calculation is a criterion-based measure that expects a particular level of student learning growth. This expected level of student learning growth takes the form of expected scale score growth on the common statewide assessment vertical scale for non-proficient (Level 1 and Level 2) students and a level maintenance expectation for students scoring proficient (Level 3 or higher) in the prior year. Unlike the value-added model, which calculates normative expectations for teacher performance based upon the performance of all students in a particular grade level in the

state for a specific year, the learning gains calculation is not associated with what actually occurred in the state in a specific year. Additionally, the learning gains expectation for Level 1 and Level 2 students is more than one year's growth, but students in Levels 3, 4, and 5 may receive learning gains even if their relative performance declines markedly.

These differences are likely to account for not only a large amount of the slippage between the two measures but also for a considerable amount of confusion in the concurrent use of these measures. The learning gains measure is more intelligible and amenable to goal-setting at the beginning of the year, but it sets lower expectations of student groups with large proportions of proficient students. The value-added measure sets even more expectations of all teachers based directly on student performance in the prior year. However, it sacrifices ease of use and understanding, particularly at the beginning of the school year.

A knowledge gap often exists among educators regarding how students are assessed and how student learning growth should be used to evaluate and develop teachers. Thus, it is critical for school leaders to know the importance of effective reading instruction and what it looks like in the classroom so that they can provide feedback to secondary reading teachers. For this to be effective, school leaders should acquire the skills and knowledge to communicate accurately with their teachers. Hattie (2009) concluded that educational leaders have a significant impact on student outcomes. Furthermore, understanding research based strategies, such as reciprocal teaching, coupled with in-depth knowledge of the curriculum yields great returns with student learning (Hattie, 2009). Therefore, in addition to classroom observation training, school

administrators in the LUSD should receive high quality professional learning with research based instructional strategies along with in-depth understanding of the content to be assessed on the standardized assessments that are used to calculate value-added measures and learning gains.

Research Question 2

What factors do middle, and high school principals and assistant principals believe contribute to the relationships among the instructional practice ratings, the value-added measures, and the learning gains?

Research Question 2 addressed LUSD school leaders' perceptions of the relationship among instructional practice scores, value-added measures, and learning gains for secondary reading teachers. Participants were able to respond to the constructed statements on the survey using a five-point Likert-type scale.

The first section of the analysis focused on the alignment of instructional practice observations with value-added scores and learning gains. The analysis revealed that approximately 44% of the participants believed that administrative observations were aligned with the value-added scores obtained by secondary reading teachers. However, not one participant strongly agreed. On the other end of the spectrum, 42% did not agree with an alignment, and 13% indicated they did not know. Reflecting on the data obtained in response to Research Question 1, there was no significant relationship between the instructional practice scores assigned to secondary reading teachers by school administrators and the value-added measure earned by the teacher. For the LUSD, the

division in perceptions is an indicator that much work remains in developing school leaders. Equally concerning is the percentage of administrators that are reluctant to commit their perspective regarding the alignment by indicating that they Did Not Know. Continued areas of focus include building capacity among administrators regarding instructional practice observations and augmenting levels of understanding regarding factors dictating value-added scores. Gordon et al. (2001) supported the idea that this LUSD, along with all school districts, should conduct comprehension checks on school administrators regarding their skill at identifying quality instruction and comparing it to actual data supporting student learning.

Florida principals and assistant principals are familiar with the concept of learning gains. Since the beginning of FCAT, learning gains have been part of the academic language of Florida school systems (Florida State Statute 1008.22, 2012) For the LUSD, the analysis of the instructional practice scores and learning gains revealed that the majority of principals and assistant principals believed there was an alignment. Despite learning gains being embedded in the culture of Florida schools, 39% of school leaders did not agree that an alignment existed between instructional practice observations and learning gains, and not one participant strongly agreed with the alignment. These findings are somewhat supportive of those of Torff and Sessions (2005) who concluded after their study that principals were more concerned about rapport with students, lesson implementation skills, and classroom management than pedagogical or content knowledge.

A large proportion of administrators in the present study did not believe that there was alignment between instructional practice scores and either of the two measures of student learning growth. The survey questions in this study were not able to provide evidence as to whether this was due to (a) a lack of confidence in the relatively new instructional practice and value-added measures or (b) a concern that though individual measures were valid they did not align well with each other. Administrators may be communicating that even though the measures are not well-aligned, they are still important to use together for a global understanding of teacher performance. This may be supported by the large percentage of administrators who stated that there was a place for all components of teacher evaluation described in this study even as they were unsure of the relationship between the measures. Another explanation, however, might be that administrators' opinions reflected a level of commitment to state requirements simply because the instructional practice and value-added measures have been required by statute.

Even though large percentages of administrators in the LUSD did not perceive an alignment to exist between instructional practice observations and value-added measures or instructional practice observations and learning gains, an overwhelming 79% of school leaders expressed the belief that the instructional practice score should be part of the summative evaluation for secondary reading teachers. According to Jacob and Lefgren (2007), school principals feel extremely capable of judging very poor as well as high quality instruction. In their study, these authors revealed that it was the 80% of instruction in the middle that challenges principals' abilities. For the LUSD, the strong

Pearson r of .48 among value-added measures and learning gains for secondary reading should be emphasized. Action steps to be taken should include professional learning of school leaders with the district observation instrument as well as developing content knowledge for the standards to be assessed for teacher value-added scores and student learning gains. Further training on the strengths and weaknesses of learning gains and value-added measures would also be helpful to administrators. Such professional learning that would permit viewing the measures in tandem rather than looking for alignment for all teachers would be more reasonable given the different assumptions in the two measures.

The last two survey items focused on instructional practice scores asked participants to gauge whether or not instructional practice scores were more reflective of overall performance than value-added or learning gains. Regarding the value-added scores, the majority of school administrators agreed that instructional practice scores were more indicative of teacher value. However, a significant percentage of respondents either disagreed or indicated they did not know. Conversely, with respect to instructional practice scores being more reflective of overall performance than learning gains, a majority of participants believed that learning gains were a more accurate data element. Based on these findings, the LUSD can be assured that the majority of its secondary administrators are comfortable in interpreting learning gains. The responses about learning gains indicate that administrators are not hesitant to use assessment-related measures of student learning growth in teacher evaluations.

Survey participants were also asked to respond to statements concerning value-added scores and their alignment to learning gains and instructional practice scores. Although a large percentage of participants believed that value-added scores and learning gains were aligned, an almost equal percentage either disagreed, strongly disagreed, or did not know. This disparity in perceptions could be explained, in part, due to the differences in the two measures which, though generally understood, have not been widely shared. There is an uneasiness among educators when it comes to value-added models, partially due to the novelty. This divide is largely due to the reality that there has been very little comprehensive research across disciplines to determine teacher impact (Newton et al., 2010). Studies conducted in large urban school districts such as Houston have shown fluctuation in teachers' value-added scores from year to year depending on their assignments (Holloway-Libell et al., 2012). The same analysis holds true for the perceptions of participants as to the alignment of value-added measures and instructional practice scores. Administrators for the LUSD were divided in their agreement and disagreement as to alignment.

Interestingly, even though the school administrators were divided, a majority expressed the belief that value-added scores should be part of the summative evaluation for secondary reading teachers. However, a majority of participants indicated that instructional practice scores and learning gains were more reflective of overall performance than value-added scores. This finding may be related to the statutory requirement that value-added scores must be used in the summative evaluation.

For the section of the survey that addressed learning gains, the theme from previous sections continued to emerge. A majority of school leaders perceived learning gains were in alignment with instructional practice scores. In addition, 76% of the principals and assistant principals believed that learning gains should be used in the summative evaluations of secondary reading teachers. Reflecting on Research Question 1, the LUSD will have to educate school leaders as well as teachers on the strong linear relationship between learning gains and value-added scores. Likewise, as previously stated, additional training is required for school leaders to align instructional practice observations against applicable standards as measured by value-added and learning gains on the state standardized assessment.

Research Question 3

To what extent do principals report using the instructional practices score, learning gains, or VAM scores to make personnel or instructional decisions?

Over 55% of administrators reported use of the instructional practice scores in making teacher assignments. A large majority stated that these scores targeted professional learning needs and yielded valuable data for meaningful conversations with teachers. Because the instructional practice scores are created through the evaluation of multiple elements of effective teaching, the results of the evaluation can be easily translated into recommendations for additional development. Upon the conclusion of the evaluation, both teachers and administrators have an understanding of areas of strength and weakness.

Despite their having had only three years of experience with the new instructional practice system, administrators communicated that they were somewhat confident in the ability of other administrators and very confident in their own ability to conduct these evaluations. Though approximately 40% of administrators stated that evaluations did not vary based upon the evaluator conducting the evaluation, slightly more (47%) disagreed. A majority of administrators believed that evaluations either differed and 13% stated that they were not sure if they differed among administrators. The large percentage that Did Not Know could indicate that administrators are conducting classroom observations in isolation. This concern did not extend to their own evaluations, however, as over 97% of administrators expressed confidence in their own ability to conduct instructional practice evaluations.

These results may indicate that further training to increase inter-rater reliability may not be helpful. Administrators' confidence in their own abilities without feelings of alignment with their colleagues provides a difficult environment in which to make changes that will increase rater reliability.

Administrators did not report either confidence in value-added measures or comfort with their use in professional learning and decision-making. A majority of administrators stated that the scores were not used in making teacher assignments, were not what was expected, were not helpful in making professional learning decisions, and were not fair. The lack of use in school decision-making can be attributed, in part, to the timing of the release of these scores. Value-added scores have been released to schools in October of the year following when the scores were earned.

Strong concerns from administrators about the fairness and relevance of these scores may be critical for teacher evaluation. A majority of administrators reported that the scores were neither what they expected nor were fair. Some of this concern may likely have arisen as a result of the manner in which value-added scores were determined over a three year period. An overwhelming majority of teachers did not receive value-added scores connected to the students they instructed in the subject area they instructed. Most teachers received “school scores” that reflected the reading and mathematics scores of students in Grades 4 through 10. Administrator opinions of fairness and relevance are likely to reflect concerns over the use of this system for teachers. Despite these deep concerns, administrators did report that the value-added data were valuable for communication with teachers. This stance may represent a form of the administrators’ commitment to and acceptance of the use of scores that have been mandated for both teacher and administrator evaluation.

Administrator confidence and comfort in the use of learning gains was a consistent theme throughout this study and held true for the use of learning gains in school decision-making. Administrators were overwhelmingly supportive of the use of learning gains in all major aspects of school planning and practice. Due to the statewide assessments release schedule, learning gains have been made available to administrators near the conclusion of the school year. These data are then available for both end-of-year conversations with teachers and other planning decisions that directly influence the coming school year. The ease of understanding and calculation of these measures may also contribute to their successful integration into these processes. Administrators

experiencing the ease of computation and comfort of use can disaggregate learning gains by class period or subgroups that provide additional information for conversations with teachers.

Qualitative Themes

Administrator free responses concerning the use of instructional practice, learning gains, and value-added information in evaluation displayed a clear separation between the three measures. Though their survey responses displayed a desire to use all three information sources for teacher evaluation and school decision-making, administrators' qualitative responses provided a more nuanced picture.

Administrators were confident in the use of instructional practice data as expressed by their free responses. Though they expressed concern with the lack of full implementation of the system, school leaders recognize a place for the use of the scores. The LUSD chose to implement the elements of the instructional practice system over three years rather than over a single year. Administrators did not express frustration over this decision but stated that their confidence that the system would increase with full implementation. Some administrators expressed concern with the differences in scoring, though these concerns were not expressed as important enough to reduce their use of the scores.

Despite other changes that have been significant over the three year implementation period, e.g., the large increase in time and training needed to implement a new instructional practice system, administrators did not mention these concerns in their

responses. This suggests that though the transition to the new system created concerns, administrators accepted these investments as worthwhile, given the data provided from the new system. Administrators have incorporated the new instructional practice evaluations into their expectations.

Though both the instructional practice and value-added evaluations were initiated in the 2011-12 school year, administrators' responses to the two different measures were very different. Administrators' responses for the instructional practice scores showed some concerns but were also hopeful for future improvement. Responses on the use of value-added scores were considerably different. A sizable portion of administrators were not only concerned about full implementation, but were either dismissive or directly ideologically opposed to the use of value-added measures. Less severe were responses about the complexity of the system and administrator concerns about the ability of teachers and administrators to understand the calculations.

No administrators commented on the methodological sophistication or other benefits of the statewide value-added model. It is unclear if this information or other information about the benefits of using a covariate adjustment model for evaluating teachers over the instructional practice and learning gains measures is widely known by administrators. This information could potentially temper frustration if the benefits of value-added modeling were more widely known. Other administrator concerns, however, such as the untimely release of score, are more difficult to change. Nearly all administrators commenting on the value-added scores stated that apart from fairness or usefulness they did not see the timeline for the score release corresponding with their

needs. The LUSD may wish to pursue an action step where the administrators are guided through the use of the value added scores when they are released. Though they are not provided soon enough to impact many school decisions, a structured and timely release of the scores with professional learning may provide opportunities to integrate the use of the scores more directly.

Implications and Recommendations for Practice

Over the last three years, the state of Florida has implemented dramatic changes in its teacher evaluation system. Professional learning has been continuous for teachers and administrators, particularly on the instructional practice system. Statewide data systems provide information on learning gains and value-added scores that make them easier to calculate and use, though the learning gains measures are more developed. Implications for practice for the school districts are made, understanding the difficult and ever-changing environment in which these evaluation changes are occurring.

For the instructional practice portion of the evaluation, it is recommended that state departments of education and local school districts continue to dedicate professional learning time and resources to develop areas of need regarding evaluation systems reliant of teacher performance. The overall inflation of the instructional practice scores suggests that administrators are facing difficulties in both understanding how elements of teaching should be scored as well as accurately recording low scores when they are appropriate. With the full implementation of the instructional practice system, school districts have an

opportunity to provide administrators with development that will assist them in increasing the rigor of this portion of the evaluation system.

Though implemented at the same time, the value-added scores have been on a different trajectory. Because many school districts limited its implementation of direct value-added measures to teachers of reading and mathematics, there was a lack of connecting applicable scores to large numbers of teachers. This resulted in school districts being very cautious in their use and support of value-added measures. When combined with the model's complexity, this has created conditions where most administrators are either unsure or skeptical of the use of the model for improving instruction and evaluating teachers.

To improve the use of value-added and other related student learning growth measures, state departments for education and local school districts should move forward with aligning all related measures to the students and courses directly instructed by teachers. Changes in state statutes have also led school districts to accelerate its movement in this direction. Direct support from the department of education and local school districts in using the value-added measures as a component of school decision-making would also be helpful. This, however, is unlikely to be enough to change embedded negative opinions about the use of value-added data. Administrators' feedback suggests that negative perceptions about the use of value-added modeling for use in high-stakes evaluation decisions and even to inform lower-stakes decisions will require levels of evidence and practical applicability far beyond that which has been provided so far. The complexity of the models and their ability to control for many

different student and school covariates has not been enough to overcome numerous other deficiencies.

The learning gains calculations did not produce many concerns from administrators, and administrators enthusiastically used this data for making decisions. Similar to value-added scores, learning gains have only been available for teachers in Grades 4 through 10, but the length of time administrators have used these scores has helped in their use. Additionally, their ease of calculation and availability at the end of the year has contributed to their relevancy. School districts should provide development and training to ensure that in the transition to the new Florida Standards Assessment that these positive uses are not lost.

A synthesis of the implications of this study for school districts, state, and national leadership include the following:

1. Increased professional learning opportunities should be provided for understanding the elements of the teacher evaluation system and gaining practice in giving feedback pertaining to low scored elements.
2. Professional learning for school administrators should include aligning observed instructional practice with applicable standards assessed on the state assessment that derive data for value-added scores and learning gains.
3. Professional learning for school administrators should include continuous updating to require inter-rater reliability training and evidence of proficiency annually.

4. School districts should move forward with aligning the value-added scores with each specific course. This will increase the level of confidence of school administrators have in value-added scores over time.
5. Emphasis should be placed on increased education of school administrators on how to utilize value-added scores as a data element for making decisions. Included in this learning opportunity should be the relationship that exists between value-added scores and learning gains.
6. School districts should work with the state department of education to deliver value-added scores strategically for school administrators so that these data can be used to assist with school based decisions as well as improving reading instruction.

Recommendations for Future Research

The expansion of teacher evaluation systems over the last decade has provided numerous opportunities for research. In this study, the researcher identified multiple areas where future research should assist in understanding and improving teacher evaluation programs.

No connection was found between instructional practice ratings and either of the quantitative measures of student learning growth, and it remains unclear if this was due to partial adoption of the instructional practice model. As school districts move toward full and mature implementation, future research should be conducted to examine how the relationships between different aspects of the evaluation system change over time among

various content areas such as mathematics and science courses. Because this study was limited to secondary schools, it is recommended that it be replicated using the same three research questions for elementary reading teachers and applicable administrators. Furthermore, a similar study should be considered that emphasizes whether significant differences exist in the perceptions of middle and high school administrators regarding the three measures: instructional practice, value-added scores, and learning gains.

Though the quantitative measures of student learning growth were found to have a moderate to strong correlation, this relationship is likely to be much stronger among groups of teachers for whom calculations of such growth are similar, e.g., teachers of Level 1 and Level 2 students. Future research should be considered to examine the data for particular groups of teachers for which value-added and learning gains measures are not aligned.

The responses for the instructional practice portion of the evaluation suggested that administrators have confidence in the instructional practice system despite this portion being relatively new. Administrators reported being confident in their ability to conduct effective evaluations but doubted the consistency in the evaluation process across their schools. Future research should be conducted to examine, when nearly all evaluations produce the same results, why administrators perceive differences in the instructional observation ratings among their peers. The tendency for administrators to inflate instructional practice scores has created the perception that some school leaders conduct evaluations incorrectly. This research should produce steps for corrective action,

such as inter-rater reliability training, that could increase confidence in the use of the evaluation system.

Administrators in this study were divided in their perceptions of whether or not instructional practice ratings were aligned with other evaluation measures. Additional studies should be conducted to clarify whether administrators view this as a concern or see this measure as a different but important aspect of instruction.

Administrators were similarly divided in their perceptions regarding the learning gains measure. Though viewed by most administrators as the strongest measure of determining teacher effectiveness, there was still a lack of confidence in the alignment of learning gains with other measures. This topic could be investigated in further detail in future studies.

The value-added measure raised multiple questions in the study that warrant additional attention. Administrators were most dismissive of this portion of the evaluation system and provided many practical examples as to why this measure did not fit into school decision-making. Future research should target a subset of administrators who report use of value-added data to clarify how administrators use these data. This should provide ideas for increasing the use of these data in the future.

The survey and free responses also indicated that a large portion of the administrators believed that the value-added measure was “unfair” or “useless.” Because of the uneven implementation process, most persons did not receive individual value-added scores. Future studies should determine if perceptions of unfairness were related to the complexity and uncertainty in the models or specific aspects of implementation.

Also, future studies should build on this study's single school district research model to survey a representative sample of small, medium, and large districts and those in non-urban contexts to determine if differences in implementation impact the alignment, use, and satisfaction with distinct elements of the evaluation system.

A synthesis of the recommendations for future research regarding instructional practice scores, value-added scores, and learning gains includes the following:

1. Beginning in the 2014-15 school year, each teacher will earn a value-added score based on the learning growth of their students for the specific course. In addition, instructional practice scores will continue to serve as a metric on the summative evaluation. Furthermore, learning gains will still exist and likely serve as a major source for school based decision making. Therefore, it is suggested that this study be replicated with other content areas such as mathematics and science.
2. It is recommended that this study be replicated to include all three research questions for elementary reading teachers and applicable school administrators and to include all students.
3. It is recommended that a similar study should be considered that emphasizes whether significant differences exist in the perceptions of middle and high school administrators regarding the three measures.
4. Because secondary reading teachers typically have students with the similar achievement levels, it is suggested that this study be replicated to examine

data for particular groups of teachers for which value-added and learning gains measures are not aligned.

5. Because a large majority of administrators have confidence in their ability to conduct instructional practice observations, follow up research should include their perceptions as to why they lack confidence regarding the observation results of their peers.
6. Because perceptions from school leaders were often divided regarding instructional practice scores, future studies should focus on these discrepancies.
7. The discrepancies in the perceptions of the school leaders regarding learning gains, though fewer in number, should be analyzed in greater detail.
8. Due to the lack of confidence in value-added scores, a study should be conducted to include a subset of administrators who use the scores for decision making.
9. Because a large percentage of administrators viewed value-added measures as unfair, an analysis of the reasons surrounding this perception should be analyzed.
10. This study should be replicated across the state to include a random sampling of small, medium, and large school districts to determine if differences in implementation impact the alignment, the use, and the satisfaction with distinct elements of the evaluation system.

Summary

The Large Unit School District (LUSD) in this study underwent dramatic changes in its evaluation system between the 2010-11 and 2013-14 school years. This study provided additional insight into the relationship between evaluation measures in this school district, administrator opinions on the appropriateness of these measures, and the use of these measures in making school decisions. The instructional practice measure did not have a significant relationship with either of the other two quantitative measures, i.e., learning gains and value-added scores, used in evaluating teachers. Both learning gains and value-added scores, however, had a moderate and significant relationship.

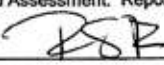
Though administrators stated that all three pieces of the evaluation system identified in this study should be used to determine educator effectiveness, they had markedly different levels of confidence and divergent opinions concerning the usefulness of the different measures. Administrators placed the highest value on the use of learning gains to determine effectiveness followed by instructional practice scores and then value-added scores. Most administrators have built both the instructional practice scores and learning gains into their decision-making processes, using them to guide professional learning and inform personnel decisions among other things. Value-added scores were an exception to this use. The late release of value-added scores and the lack of administrator confidence in their usefulness have led to limited use by administrators for decision making purposes.

As school districts implement new evaluation systems, administrator feedback is critical to making continuous improvement. Administrators do not only complete the

evaluations, they also create building level support for the implementation of the entire evaluation system. Their values and preferences impact the amount of emphasis placed on aspects of the evaluation, the fidelity of implementation, and how the results are used in the evaluation process. The feedback from administrators in the LUSD indicate that, despite clear progress in the improvement of the evaluation system, some portions, namely the value-added measures, are underdeveloped and will require additional attention.

APPENDIX A
SCHOOL DISTRICT PERMISSION TO CONDUCT THE STUDY

Submit this form and a copy of your proposal to:		RESEARCH REQUEST FORM		Your research proposal should include: <ul style="list-style-type: none"> Project Title Purpose and Research Problem Instruments Procedures and Proposed Data Analysis 	
Requester's Name <u>Ronald Scott Fritz</u>		Date <u>09/28/13</u>			
E-mail _____		Phone _____			
Address _____		City, State _____		Zip _____	
Institutional Affiliation <u>UCF</u>					
Project Director or Advisor <u>Dr. Rose Taylor</u>		Phone <u>407-399-</u>			
Degree Sought: (check one)		<input type="checkbox"/> Associate <input checked="" type="checkbox"/> Doctorate		<input type="checkbox"/> Bachelor's <input type="checkbox"/> Not Applicable	
		<input type="checkbox"/> Master's		<input type="checkbox"/> Specialist	
Project Title: <u>AN ANALYSIS OF THE DIFFERENCE BETWEEN THE ASSESSED INSTRUCTIONAL PRACTICES AND THE VALUE-ADDED MEASURE OF TEACHERS</u>					
ESTIMATED INVOLVEMENT					
PERSONNEL/CENTERS	NUMBER	AMOUNT OF TIME (DAYS, HOURS, ETC.)	SPECIFY SCHOOLS BY NAME AND NUMBER OF TEACHERS, ADMINISTRATORS, ETC.		
Students					
Teachers					
Administrators	434	1 hour	all schools and all prinipals and assistant principals		
Schools/Centers					
Others (specify)					
Specify possible benefits to students/school system: <u>Determine the alignment of value-added measures assigned and the instructional practice observations performed by school based administrators.</u>					

ASSURANCE					
Using the proposed procedures and instrument, I hereby agree to conduct research in accordance with the policies of the Orange County Public Schools. Deviations from the approved procedures shall be cleared through the Senior Director of Accountability, Research, and Assessment. Reports and materials shall be supplied as specified.					
Requester's Signature <u></u>					
Approval Granted: <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No		Date: <u>October 2nd, 2013</u>			
Signature of the Senior Director for Accountability, Research, and Assessment _____					

APPENDIX B
UNIVERSITY OF CENTRAL FLORIDA
INSTITUTIONAL REVIEW BOARD APPROVAL



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

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

To: **Ronald Fritz**

Date: **November 27, 2013**

Dear Researcher:

On 11/27/2013, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	AN ANALYSIS OF THE DIFFERENCE BETWEEN THE ASSESSED INSTRUCTIONAL PRACTICES, VALUE-ADDED MEASURE, AND LEARNING GAINS OF TEACHERS.
Investigator:	Ronald Fritz
IRB Number:	SBE-13-09735
Funding Agency:	
Grant Title:	
Research ID:	na

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

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 11/27/2013 03:56:49 PM EST

IRB Coordinator

APPENDIX C
COMMUNICATIONS WITH PARTICIPANTS

Initial Participant Communication

Thank you for taking the time to read this email. I invite you to complete a short survey that was created as part of my doctoral study designed to address the following research questions: (a) What factors do middle and high school principals and assistant principals for instruction believe contribute to the relationships among the instructional practice rating value-added measures, and learning gains? (b) To what extent do principals report using the instructional practice score, learning gains, or VAM scores to make personnel or instructional decisions?

The survey will only take a few minutes to complete, and the Educational Leadership Executive Ed.D., Program Coordinator and faculty have approved this study. There are no perceived benefits, or anticipated risks for participating in this study as your identity and responses are confidential. Your participation, though encouraged, is voluntary and you may decline to participate at any time without penalty. Also, you do not have to answer any questions that you do not wish to. Data and results will be analyzed and reported in aggregate form, not by individual student response or demographic information. Your name, and any other identifiable information will not be associated with responses.

Thank you in advance for your participation in this study. Please complete this survey by Tuesday, April 8, 2014. Your responses will be valuable in the continual improvement of the Ed.D. in Educational Leadership Executive program. To complete the survey please click on the following link.

Follow this link to the Survey:

Take the

Survey<http://ucf.qualtrics.com//WRQualtricsSurveyEngine/?Q_SS=5b9PxW0XhhTgceh_e5rYxZGeOYrNH1z&_=1>

Or copy and paste the URL below into your internet browser:

http://ucf.qualtrics.com//WRQualtricsSurveyEngine/?Q_SS=5b9PxW0XhhTgceh_e5rYxZGeOYrNH1z&_=1

Scott Fritz

Dear Participant,

If you have completed and submitted the survey regarding Instructional Practice Observations, Value-Added measures, and Learning Gains, I sincerely appreciate your efforts. If you still need to complete the survey, please follow the link below. It would be appreciated if you would complete the survey by Wednesday May 14th. Your response is extremely important as it will help guide the work of our district as well as promote the best practices for administrators throughout the state.

If you have questions, my personal cell is xxx-xxx-xxxx. Thank you for your participation.

Scott Fritz

Dear STARTED,

Thank you for starting the survey regarding Instructional Practice Observations, Value-Added measures, and Learning Gains. Your response is extremely important as it will help guide the work of our district as well as promote the best practices for administrators throughout the state.

Please follow the link below and complete today. If you have questions, my personal cell is xxx-xxx-xxxx. Thank you for your participation.

Scott Fritz

Dear NOT STARTED,

This is a reminder that I need you to complete the survey regarding Instructional Practice Observations, Value-Added measures, and Learning Gains. Your response is extremely important as it will help guide the work of our district as well as promote the best practices for administrators throughout the state.

Please follow the link below and complete today. If you have questions, my personal cell is xxx-xxx-xxxx. Thank you for your participation.

Scott Fritz

APPENDIX D
INSTRUCTIONAL PRACTICE, VALUE-ADDED MEASURE,
SECONDARY READING LEARNING GAINS SURVEY



Section 1: School Demographic Information

Please indicate the demographic information for your school during the 2012-2013 school year.

1. Percent of students receiving free or reduced lunch benefits:

- ☐ Less than 40
- ☐ 41-50
- ☐ 51-60
- ☐ 61-70
- ☐ 71-80
- ☐ 81-100

2. Percent of students receiving special education services:

- ☐ Less than 5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ 21-25
- ☐ 26 or greater

3. Percent of students identified as English language learners:

- ☐ Less than 5
- ☐ 5 - 9.9
- ☐ 10 - 19.9
- ☐ 20 - 29.9
- ☐ Higher than 30

4. Years of experience in current position:

- ☐ 0-3
- ☐ 3-5
- ☐ 5-10
- ☐ 11-15
- ☐ more than 15

5. Total years of experience as an administrator:

- ☐ 0-3
- ☐ 3-5
- ☐ 5-9
- ☐ 10-15
- ☐ more than 15

6. Highest degree earned:

- ☐ Master's
- ☐ Education Specialist
- ☐ Doctorate

7. Gender:

- ☐ Male
- ☐ Female

8. Type of school:

- ☐ Traditional
- ☐ K-8
- ☐ Exceptional Education Center
- ☐ Alternative
- ☐ Charter

Section 2: Instructional Practice Observations

As the instructional leader, reflect on how you have used the 2012-2013 summative Instructional Practice scores during the 2013-2014 school year for the reading teachers in your current school. Please respond to the following statements.

9. For the 2012-2013 school year, the instructional practice scores given to reading teachers were aligned with the learning gains of their students.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

10. For the 2012-2013 school year, the instructional practice scores given to reading teachers were aligned with their assigned value-added measures.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

11. For the 2012-2013 school year, the instructional practice scores given to the reading teachers should be used as a portion of their summative performance evaluations.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

12. For the 2012-2013 school year, the instructional practice scores given to the reading teachers were used in making teaching assignments for the 2013 - 2014 school year.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

13. For the 2012-2013 school year, the instructional practice scores assigned to the reading teachers more accurately reflected their overall performance, than the value-added measure they were assigned.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

14. For the 2012-2013 school year, the instructional practice scores earned by the reading teachers more accurately reflected their overall performance than the learning gains achieved by their students.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree

- ☐ Strongly Agree
- ☐ Do not know

15. For the 2012-2013 school year, the instructional practice scores earned by the reading teachers during a single observation varied, depending on which administrator conducted the observation.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

16. For the 2012-2013 school year, the instructional practice scores earned by the reading teachers helped you target their professional learning needs for 2013 - 2014 school year.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

17. For the 2012-2013 school year, the instructional practice scores assigned to the reading teachers gave you valuable data in order to have meaningful discussions regarding their improvement in effective instructional delivery.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

18. You are confident that you have the expertise in order to conduct instructional practice observations.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

Section 3: Value-Added Measure

As the instructional leader, reflect on how you have used the 2012-2013 Value-Added scores during the 2013-2014 school year for the reading teachers in your current school. Please respond to the following statements.

19. For the 2012-2013 school year, the value-added measures assigned to the reading teachers aligned with the learning gains of their students.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

20. For the 2012-2013 school year, the value-added measures assigned to the reading teachers aligned with their summative instructional practice scores.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

21. For the 2012-2013 school year, the value-added measures assigned to the reading teachers should be used as a portion of their summative performance evaluations.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

22. For the 2012-2013 school year, the value-added measures assigned to the reading teachers were a determinate of how staff were assigned for 2013-2014.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

23. For the 2012-2013 school year, the value-added measures assigned to the reading teachers more accurately reflected their overall performance than the summative instructional practice scores.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

24. For the 2012-2013 school year, the value-added measures assigned to the reading teachers more accurately reflected their overall performance than the learning gains earned by their students.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

25. For the 2012-2013 school year, the value-added measures assigned to the reading teachers were what you expected.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

26. For the 2012-2013 school year, the value-added measures assigned to the reading teachers helped you target their professional learning needs for the 2013-2014 school year.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

27. For the 2012-2013 school year, the value-added measure assigned to the reading teachers gave you valuable data in order to have meaningful discussions regarding their improvement in effective instructional delivery.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree

☐ Do not know

28. For the 2012-2013 school year, I believe that the value-added measures assigned to the reading teachers were fair.

- ☐ Strongly Disagree
☐ Disagree
☐ Agree
☐ Strongly Agree
☐ Do not know

Section 4: Learning Gains

As the instructional leader, reflect on how you have used the 2012-2013 Learning Gains during the 2013-2014 school year for the reading teachers in your current school. Please respond to the following statements.

29. For the 2012-2013 school year, the learning gains for the reading students in this school reflect the quality of the reading instruction.

- ☐ Strongly Disagree
☐ Disagree
☐ Agree
☐ Strongly Agree
☐ Do Not Know

30. For the 2012-2013 school year, the learning gains for the reading teachers reflect their instructional practice scores.

- ☐ Strongly Disagree
☐ Disagree
☐ Agree
☐ Strongly Agree
☐ Do not know

31. For the 2012-2013 school year, the learning gains for the reading teachers should be used as a portion of their summative performance evaluations.

- ☐ Strongly Disagree

- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

32 For the 2012-2013 school year, the learning gains for the reading teachers were used in making decisions regarding teaching assignments for the 2013-2014 school year.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

33 For the 2012-2013 school year, the learning gains for the reading teachers helped you target the professional learning needs of reading teachers for the 2013-2014 school year.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

34 For the 2012-2013 school year, the learning gains assigned to the reading teachers gave you valuable data in order to have meaningful discussions regarding their improvement in effective instructional delivery.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree
- ☐ Do not know

35 For the 2012-2013 school year, the learning gains for the reading teachers prompted you to review the scope and sequence utilized in reading classes.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Agree
- ☐ Strongly Agree

☐ Do not know

36. For the 2012-2013 school year, the learning gains for the reading teachers prompted you to review the instructional materials utilized in reading classes.

- ☐ Strongly Disagree
☐ Disagree
☐ Agree
☐ Strongly Agree
☐ Do not know

37. For the 2012-2013 school year, the learning gains for the reading teachers prompted you to review how class time was utilized in reading classes.

- ☐ Strongly Disagree
☐ Disagree
☐ Agree
☐ Strongly Agree
☐ Do not know

Section 5: Please respond to the open ended items:

38. As the instructional leader, reflect on the 2012-2013 learning gains, value-added measures, and instructional practice scores. To what extent do you use this data to improve the reading performance at your school for the 2013-2014 school year.

39. Please share with the researcher anything that you believe would be helpful in improving the effectiveness of reading teachers.

Qualtrics Survey Software

Comments/Suggestions



© University of Central Florida
4000 Central Florida Blvd., Orlando, FL 32816 | 407.823.2000

Survey Powered By [Qualtrics](#)

<https://ucf.qualtrics.com/ControlPanel/Display.asp?surveyid=Service&T=66mU3G90014216297164>

REFERENCES

- Alexander, K., & Alexander, M. D. (2012). *American public school law* (8th ed. ed.). Belmont, CA: Wadsworth Cengage Learning.
- American Institutes For Research (2010). *Value-added model for measuring student learning gains: General approach and framework for evaluating candidate models*. Retrieved from <http://www.fldoe.org/committees/pdf/sghandout.pdf>
- American Institutes for Research. (2011, August 1-2). *Florida's value-added technical assistance workshop* [PowerPoint]. Retrieved from <http://www.fldoe.org/Committees/pdf/august12tampres.pdf>
- American Recovery and Reinvestment Act of 2009, Pub. L No. 1-407, §.
- Amerein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS EVAAS) in the Houston independent school district (HISD): intended and unintended consequences. *Educational Policy Archives*, 20(12). Retrieved from: <http://epaa.asu.edu/ojs/article/view/1096>
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessments of teachers. *Journal of Educational and Behavioral Statistics* 29(37) DOI: 10.3102/10769986029001037
- Beteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping Stones: Principal Career Paths and School Outcomes. *Social Science Research*, 41(4), 904–919.
- Biancarosa, G., Bryk, A. B, & Dexter, E. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal*, 111, 7-34.

- Bill & Melinda Gates Foundation (2010). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Retrieved from:
http://www.metproject.org/downloads/MET_Gathering_Feedback_Practitioner_Brief.pdf
- Bill & Melinda Gates Foundation. (2013). Retrieved from:
http://www.metproject.org/downloads/MET_Gathering_Feedback_Practitioner_Brief.pdf
- Briggs, D. C. (2011). *Making value-added inferences from large-scale assessments*. Retrieved from:
http://www.colorado.edu/education/faculty/derekbriggs/Docs/Briggs_VAM%20Inferences_101511.pdf
- Byrk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Sage.
- Chin, T. Y., Shaw, L. H., Dwyer, A. C., McCormick, C., & Geisinger, K. F. (2010). *Operational check of the 2010 FCAT 10th grade reading and 11th grade science equating results*. Retrieved from:
<http://www.fldoe.org/pdf/2010fcatcalibrationequatinggr10-11.pdf>
- Cincinnati Public Schools Collective Bargaining Agreement (n.d.). Retrieved from:
http://www.gse.harvard.edu/~ngt/par/resources/contracts_cincinnati.pdf
- Collins, C., & Amrein-Beardsley, A. (2012). Putting growth and value-added models on the map: a national overview. *Teachers College Record*, 116(1), 2014, xx-xx.

- Connor, W. (1920). A new method for rating teachers. *The Journal of Educational Research* 1(5), 338-358. Retrieved from: <http://www.jstor.org/stable/27524324>
- Cornett, L., & Bailey, A. A. (2003). *Focus on “high qualified” teachers: SREB stats take action*. Retrieved from http://www/sreb/org/scripts/Focus/Reports/Highly_qualified.pdf
- Danielson, C. (2011). *Evaluations that help teachers learn*. Retrieved from: <http://www.ascd.org/publications/educational-leadership/dec10/vol68/num04/Evaluations-That-Help-Teachers-Learn.aspx>
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (1996). What matters most? A competent teacher for every child. *Phi Delta Kappan*, 78(3), 193-200.
- Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education*, 51(3), 166-173.
- Darling-Hammond, L. (2009). America’s commitment to equity will determine our future. *Phi Delta Kappan* 91(4), 8-14.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Dilman, D. A, Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York, NY: Wiley.

- Donaldson, G. (1993). Working smarter together: Leading collaborative change in schools means helping staffs become more productive without substantially depleting their resources. *Educational Leadership* 51(2), 12-16.
- Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Washington, DC: Center for American Progress.
- Fink, S. (2012). *School and district leaders as instructional experts: What we are learning*. Seattle, WA: Center for Educational Leadership, University of Washington.
- Finn, C. (1999). *Better teachers, better schools*. Washington, DC: Fordham Foundation.
- Florida Department of Education (2013a). *Cognitive complexity of FCAT test items*. Retrieved from: http://fcat.fldoe.org/pdf/cog_complexity-fv31.pdf
- Florida Department of Education (2013b). Retrieved from: <http://fcat.fldoe.org>
- Florida Department of Education (2013c). Retrieved from: <http://www.fldoe.org/profdev/pdf/teacherevaluationresults.pdf>
- Florida Department of Education (2013d). Retrieved from <http://schoolgrades.fldoe.org/pdf/1112/SchoolGradesTAP2012.pdf>
- Florida Department of Education (2014a). FCAT Results for Orange County. Retrieved from: <http://fcat.fldoe.org/resultsFCAT2/default.asp>
- Florida Department of Education (2014b). Florida Value-Added Model Technical Report. Retrieved from: <http://www.fldoe.org/.../Value-Added-Model-Technical-Reports.docx>

Florida Department of Education (2014c). Presentation Value-Added Model. Retrieved

from: <http://www.fldoe.org/committees/pdf/PresentationValue-addedModel.pdf>

Florida Department of Education (2014d). Value-Added-Model White-Paper. Retrieved

from: www.fldoe.org/committees/.../Value-Added-Model-White-Paper.doc

Florida Organization of Instructional Leaders (2012). Retrieved from

<http://www.floridafoil.com/wp-content/uploads/2011/08/Coalition-Powerpoint.ppt-April-9-20121.pdf>

Florida State Statute 1008.22 (2012), Retrieved from

http://leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=1000-1099/1008/Sections/1008.22.html

Florida State Statute 1012.28 (201), Retrieved from

<http://www.flsenate.gov/laws/statutes/2011/1012.28>

Florida State Statute 1012.34 (2012). Retrieved from

http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=1000-1099/1012/Sections/1012.34.html

Fraenkel, J. R., Wallen, N., E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. New York, NY: McGraw-Hill.

Friedman, B. A., Friedman, M. A., & Markow, D. (2008). Predictors of principals' satisfaction with their schools. *Journal of Educational Administration*, 46(5), 598-612. Doi: 10.1108/09578230810895519

- Gansle, K. A., Noell, G. H., & Burns, J. M. (2012) Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education*, 63, 304-317.
- Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brookings. Retrieved from: <http://www.brookings.edu/research/reports/2010/11/17-evaluating-teachers>
- Gordon, B. G., Meadows, R. B., & Dyal, A. B. (2001). School principals' perceptions: the use of formal observations of classroom teaching to improve instruction. *Education*, 116(1), 9-15.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge, Taylor, & Francis.
- Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures performance: a meta-analysis. *Personnel Psychology*, 39, 811-826.
- Hess, K. K., Carlock, D., Jones, B., & Walkup, J. R. (2009). *What exactly do "fewer, clearer, and higher standards" really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments*. Unpublished manuscript. Retrieved from http://www.nciea.org/cgi-bin/pubspage.cgi?sortby=pub_date
- Holloway-Libell, J., Armrein-Beardsley, A., & Collins, C. (2012). All hat and no cattle, *Educational Leadership* 70(3), 65-68

- Jackson, R. R. (2008). The instructional leader's guide to strategic conversations with teachers: *Mindsteps*.
- Jacob, B. A., & Lefgren, L. (2007). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
- Jerald, C., D. (2009). *The value of value-added data*. Washington, DC: The Education Trust.
- Joyce, B., & Showers, B. (2002). *Student achievement through staff development*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kane, T., & Cantrell, S. (2012). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle, WA: Bill and Melinda Gates Foundation.
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high quality observations with student surveys and achievement gains* (Policy and Practice Brief). Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf
- Kane, T., Taylor, E. S., Tyler, J. H., and Wooten, A. L. (2010). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3), 587-613.
- Kelly, S. (2004). Are teachers tracked? On what basis and with what consequences? *Social Psychology of Education*, 7, 55-72.

- Kersting, N. B., Chen, M., & Stigler, J. W. (2012). Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Educational Policy Analysis Archives* 21(7), 1-39.
- Koedinger, K., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36, 757-798. doi: 10.1111/j.1551-6709.2012.01245.x
- Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education*, 33(4), 279-299.
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value-added assessment system. *Educational Evaluation & Policy Analysis*, 25, 287-298.
- Lam, S. (2001). Educators' opinions on classroom observation as a practice of staff development and appraisal. *Teacher and Teacher Education*, 17(2), 161-73.
- Large Urban School District (2011). *Teacher evaluation system proposal manual*. Author.
- Large Urban School District (2013a.) *Instructional personnel evaluation system procedures manual 2012-13*. Retrieved from <https://www.ocps.net/es/hr/PDS/assessment/Additional%20Information/Evaluation%20Manual%202012-13%20Final.pdf>

Large Urban School District (2013b). *District report from iObservation 2012-2013*.

Large Urban School District (2014). Personnel information report. Retrieved from
Human Resource Department.

Lockwood, J., McCaffrey, D., Mariano, L., & Setodji, C. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics*, 32, 125-150.

Loeb, S., Kalogrides, D., & Beteille, T. (2012). *Effective schools: Teacher hiring, assignment, development, and retention*. Retrieved from
<http://cepa.stanford.edu/publications/school-leadership> p 269-305

Lunenburg, F. C., & Irby, B. J. (2008). *Writing a successful thesis or dissertation: Tips and strategies for students in the social and behavioral sciences*. Thousand Oaks, CA: Corwin.

Marshall, K. (2012). Fine-tuning teacher evaluation. *Educational Leadership* 70(3), 50-53.

Marzano, R. (2004). *Building background knowledge for academic achievement: Research on what works in schools*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. (2010). Art & science of teaching/What teachers gain from deliberate practice. *The Effective Educator*, 68(4), 82-85.

Marzano, R. J. , (2007). *The art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Marzano, R. J., Boogren, T., Heflebower, T., Kanold-McIntyre, J., & Pickering, D. (2012). *Becoming a reflective teacher*. Bloomington, IN: Marzano Research Laboratory.
- Marzano R. J., & Simms, J. A. (2013). *Coaching classroom instruction*. Bloomington, IN: Marzano Research Laboratory.
- Marzano, R. J., Toth, M., & Schooling, P., (2012). Retrieved from <http://ok.gov/sde/sites/ok.gov.sde/files/TLE-MarzanoWhitePaper.pdf>
- Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research*, 80(4), 242-247.
- Mertz, N. T. (2010). Teacher selection and school leader effects. *Journal of School Leadership*, 20(2), 184-207.
- Murnane, R. J., Singer J. D., Willett, J. B., Kemple, J. J., & Oden, R. J. (1991). *Who will teach? Policies that matter*. Cambridge, MA: Harvard University Press.
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: an exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(1), 1-24.
- Nixon, A., Packard, A. L., & Dam, M. (2011). *School principals and teacher contract non-renewal*. Retrieved from: <http://creativecommons.org/licenses/by/3.0/>
- Pella, S. (2012). What should count as data or data-driven instruction? Toward contextualized data-inquiry models for teacher education and professional development. *Middle Grades Research Journal*, 7(1), 57-75.

- Peterson, K. D. (2000). Research on school teacher evaluation. *NASSP Bulletin*, 88(639), 60-79.
- Prendergast, C., & Topel, R. H. (1993) 'Discretion and bias in performance evaluation.' *European Economic Review*, 37(2-3), 355-365.
- Quick, H. E., Holtzman, D. J., & Chaney, K. R. (2009). Professional development and instructional practice. Conceptions and evidence of effectiveness. *Journal of Education for Students Placed at Risk*, 14(1), 45-71., DOI: 10.1080/10824660802715429
- Ravitch, D. (2010). *The death and life of the American school system*. New York, NY: Perseus Books.
- Rockoff, J., E. (2004). The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review* 94(2), 247-252.
- Ross, S. M., Wang, L. W., Alberg, M., Sanders, W. L., Wright, S. P., & Stringfield, S. (2001, April). *Fourth-year achievement results on the Tennessee value-added assessment system for restructuring schools in Memphis*. Paper presented at the meeting of the American Educational Research Association, Seattle, WA.
- Sanders, W. L., & Horn, S. P., (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299-311.
- Sanders W. L., & Horn, S. P. (1998). *Research findings from the Tennessee value-added assessment system (TVASS) database: Implications for educational evaluation and research*. Retrieved from: http://www.sas.com/govedu/edu/ed_eval.pdf

- Sanders W. L., & Rivers J. C. (1996). *Research progress report: Cumulative and residual effects of teachers on future student academic achievement*. Retrieved from: [http://www.cgp.upenn.edu/pdf/Sanders_Rivers-TVASS teacher%20effects.pdf](http://www.cgp.upenn.edu/pdf/Sanders_Rivers-TVASS_teacher%20effects.pdf)
- Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. (Research Report). Consortium on Chicago School Research, University of Chicago Urban Education Institute. Retrieved from <http://ccsr.uchicago.edu/sites/default/files/publications/Teacher%20Eval%20Report%20FINAL.pdf>
- Sawchuk, S. (2012, October 24). 'Value-added' measures at the secondary level questioned. *Education Week*, p. 6. Retrieved from: http://www.edweek.org/ew/articles/2012/10/24/09tracking_ep.h32.html?qs=andrew_anderson
- Schatschneider, C., Buck, J., Torgesen, J., & Wagner, R. (2004). *A multivariate study of individual differences in performance on the reading portion of the Florida Comprehensive Assessment Test: A brief report*. Retrieved from: http://www.fcrr.org/technicalreports/multi_variate_study_december2004.pdf
- Shelton, S. (2010). *Strong leaders strong schools: 2010 school leadership laws*. Denver, CO: National Conference of State Legislatures. Retrieved from

- <http://www.wallacefoundation.org/knowledge-center/school-leadership/state-policy/Documents/2010-Strong-Leaders-Strong-Schools.pdf>
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1-22.
- Steinberg, W., J. (2011). *Statistics alive*. (2nd. ed.). Thousand Oaks, CA: Sage.
- Stone, J. E. (2002). *The value-added achievement gains of NBPTS-certified teachers in Tennessee: A brief report*. Arlington, VA: Education Consumers Foundation. Retrieved from http://www.education-consumers.org/research/briefs_0502.htm
- Torff, B., & Sessions, D. (2005). Principals' perceptions of the causes of teacher ineffectiveness. *Journal of Educational Psychology*, 97(4), 530-537. doi: 10.1037/0022-0663.97.4.530
- Torgesen, J., Nettles, S., Howard, P., & Winterbottom, R. (2004). *Brief report of a study to investigate the relationship between several brief measures of reading fluency and performance on the Florida Comprehensive Assessment Test-Reading in 4th, 6th, 8th, 10th grades*. Retrieved from: http://www.fcrr.org/TechnicalReports/Progress_monitoring_report.pdf
- Tucker, P. (2001). Helping struggling students. *Educational Leadership*, 58(5), 52-55.
- U.S. Department of Education (2012). *Race to the top. Florida report year 1: School year 2010-2011*. Retrieved from <http://www2.ed.gov/programs/racetothetop/performance/florida-year-1.pdf>

- U.S. Department of Education (2013). *Race to the top. Florida report year 2: School year 2011-2012*. Retrieved from
<http://www2.ed.gov/programs/racetothetop/performance/florida-year-2.pdf>
- Wallace Foundation (2013). *The school principal as leader: Guiding schools to better teaching and learning*. New York, NY: Author. Retrieved from
<http://www.wallacefoundation.org/knowledge-center/school-leadership/effective-principal-leadership/Pages/The-School-Principal-as-Leader-Guiding-Schools-to-Better-Teaching-and-Learning.aspx>
- Watkins, P. (2005). The principal's role in attracting, retaining, and developing new teachers: Three strategies for collaboration and support. *Clearing House*, 79(2), 83-87.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on teacher effectiveness*. New York, NY: The New Teacher Project.
- Winters, M., A., (2012). *Transforming tenure: Using value-added modeling to identify ineffective teachers*. (Civic Report No. 70). Retrieved from:
http://www.manhattan-institute.org/pdf/cr_70.pdf