
Electronic Theses and Dissertations, 2004-2019

2016

Content and Effects of Specific Targeted Feedback from Teacher Observations on Student Achievement in a Large Urban School District.

Rachel Larsen
University of Central Florida



Part of the [Education 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, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Larsen, Rachel, "Content and Effects of Specific Targeted Feedback from Teacher Observations on Student Achievement in a Large Urban School District." (2016). *Electronic Theses and Dissertations, 2004-2019*. 5055.

<https://stars.library.ucf.edu/etd/5055>

CONTENT AND EFFECTS OF SPECIFIC TARGETED FEEDBACK FROM
TEACHER OBSERVATIONS ON STUDENT ACHIEVEMENT IN A LARGE
URBAN SCHOOL DISTRICT

by

RACHEL MARISSA LARSEN
B.A. Florida State University, 2006
M.A. University of Central Florida, 2009

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

Summer Term
2016

Major Professor: Barbara Murray

© 2016 Rachel M Haynes

ABSTRACT

The purpose of this study was to examine the relationships between the content of specific targeted feedback and student achievement and to identify student and teacher knowledge voids from the content of feedback in relationship to achievement level outcomes. This study aimed to deeply analyze the effects of specific targeted feedback as researched by Rafalski (2015).

De-identified data from a purposive sample of Rafalski's (2015) original study were selected from a large urban school district in Central Florida as well as student achievement level outcome data from the 2013-2014 Florida Comprehensive Assessment Test 2.0 (FCAT 2.0). Raw observation data were used to determine the content of the feedback and analyzed with statistical tests to identify if relationships existed between the feedback content and student achievement. The observation data came from the 2013-2014 Marzano protocol used in the large urban school district in this study. Data were coded from a rubric created to determine the categories and charges of specific targeted feedback. Descriptive statistics were calculated to identify frequencies in the data, and Pearson's r was used to calculate correlations between the categories and charges of feedback coded and the average student achievement level outcomes for FCAT 2.0 assessments for each teacher in the study.

Data from frequency calculations showed areas in which elements scored, feedback categories, and feedback charge were heaviest and lacking. Correlations showed areas in which statistically significant relationships occurred and did not occur. From these data it was determined that in terms of supporting level 1 and level 2 students,

teachers did not receive much feedback in areas of building student relationships, and probing low expectancy students. Teachers also received predominantly neutral feedback and feedback that contained coaching tips. The validity and relevancy of the feedback was beyond the scope of this study. Correlational data showed both positive and negative relationships between elements coded and student achievement level outcomes as well as feedback categories and student achievement level outcomes. There were no statistically significant relationships between the charge of feedback and student achievement data.

For my daughter, Lily

ACKNOWLEDGMENTS

This dissertation would not have been possible without the constant and unquestioned support from my family. This is for my daughter, Lily; she is my heart and soul and the reason why I continue to strive for excellence. Thank you to my parents for babysitting, helping to ensure that I was able to make it to class on time, and for always instilling in me the importance of education. To my little brother, Michael, who was also working on becoming a doctor in medical school and liked to (playfully) remind me that he not only graduated three months before me, but that he is the “real” doctor in the family. Thank you to my loving husband, James, who went through this process with me, listened and celebrated with me, answered my questions and advised me, and supported my future career aspirations.

I would like to recognize my committee chair, Dr. Barbara Murray. She has been a constant support in helping me to maintain my focus, push me to the end, and show me that my work can make a difference in education. I would also like to thank the committee members who were all a part of my success: Dr. Lee Baldwin, Dr. Walter Doherty, and Dr. Lynn Spadaccini. I would like to extend a special thank you to Dr. Shana Rafalski who set the foundation for my own study, and also to the loving and wonderful Dr. Renee Goodstein, who helped me edit and polish my work. The expertise, knowledge, and guidance from these professionals helped make this experience rich and fulfilling and better prepared me for a career in educational leadership.

Throughout this process in the last three years, I have been exposed to many people who have imparted on me their knowledge of what it is like to be a leader and to

navigate the inner workings of people and education. Thank you to the many people who helped make me a better leader by teaching me, sharing experiences, giving me advice, keeping me grounded, and kept me laughing until the end. Thank you to the fantastic cohort four for making classes fun, entertaining, and for striving to stick together to graduation! Thank you to all the people at the University of Central Florida and Orange County Public Schools for answering my questions and guiding me toward the resources I needed.

This dissertation and experience has opened my eyes to the real world of education and helped me realize in what direction I want to steer my career. I have learned about being a more prepared and knowledgeable individual, and I have learned that I will never stop learning.

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xi
CHAPTER 1 PROBLEM OF PRACTICE	1
Background of the Study	1
Problem Statement	5
Purpose Statement.....	6
Significance of the Study	6
Definition of Terms.....	8
Definitions as Related to Assessment	9
Definitions as Related to Observation	10
Definitions as Related to Feedback.....	12
Definitions as Related to Professional Development.....	13
Operational Definitions.....	14
Theoretical Framework.....	15
Research Questions	21
Limitations	23
Delimitations.....	24
Assumptions.....	25
Organization of the Study	27
Methodology	27
Procedures.....	28
Population and Sample	29
Data Collection	30
Data Analysis	31
CHAPTER 2 REVIEW OF LITERATURE	33
Introduction.....	33
Accountability and Assessment	36
Evaluation	42
VAM	48
Marzano	53
Supervision	63
Feedback	68
Coaching.....	78
Professional Development	83
Summary	93
CHAPTER 3 METHODOLGY	95
Introduction.....	95
Selection of Participants	97
Instrumentation	98

Data Collection	101
Data Analysis	104
Research Question 1	107
Research Question 2	107
Research Question 3	108
Research Question 4	109
Research Question 5	110
CHAPTER 4 ANALYSIS OF DATA	115
Introduction.....	115
Research Question 1	115
Research Question 2	120
Research Question 3	132
Research Question 4	135
Research Question 5	139
Summary.....	141
CHAPTER 5 SUMMARY	150
Introduction.....	150
Summary of the Study	150
Discussion of the Findings.....	156
Implications for Practice.....	180
Recommendations for Further Research.....	189
Conclusions.....	196
APPENDIX A UNIVERISTY OF CENTRAL FLORIDA INSTITUTIONAL REVIEW BOARD APPROVAL OF RESEARCH	199
APPENDIX B LARGE URBAN SCHOOL DISTRICT INSTITUTIONAL REVIEW BOARD APPROVAL OF RESEARCH	201
APPENDIX C SAMPLE RUBRIC FROM RAFALSKI STUDY	203
APPENDIX D RUBRIC FOR SPECIFIC TARGETED FEEDBACK CATEGORIES	206
APPENDIX E CHARGE RUBRIC	208
APPENDIX F CHARGE AND CATEGORY EXAMPLE MATRIX	210
APPENDIX G EXAMPLES OF IMPROPER FEEDBACK COMMENTS	215
APPENDIX H MARZANO LEARNING MAP.....	218
REFERENCES	221

LIST OF FIGURES

Figure 1 <i>Difference of performance in Florida's Value-Added Model</i>	53
Figure 2 <i>"A model of feedback"</i>	77
Figure 3 <i>"Teacher enquiry and knowledge-building cycle to promote important outcomes for students."</i>	93
Figure 4 <i>Achievement levels for FCAT 2.0 Reading assessment grades 3 through 10</i> ..	100
Figure 5 <i>Achievement levels for FCAT 2.0 Mathematics assessment grades 3-8</i>	101

LIST OF TABLES

Table 1 Teachers by Predominant Feedback Type	98
Table 2 Research Questions, Variables, Sources of Data and Methods of Analysis	113
Table 3 Description of Student Achievement Levels on FCAT 2.0 Assessments.....	116
Table 4 Frequency of Student Achievement Levels of Teachers with Specific Targeted Feedback on Reading FCAT 2.0.....	117
Table 5 Frequency of Student Achievement Levels of Teachers with Specific Targeted Feedback on Mathematics FCAT 2.0	118
Table 6 Frequency of Student Achievement Levels of Teachers with Specific Targeted Feedback on Reading FCAT 2.0 Retake.....	119
Table 7 Frequency of Element and Design Questions Scored Receiving Specific Targeted Feedback	121
Table 8 Pearson Correlations between Elements Scored and FCAT 2.0 Mean Student Achievement Levels.....	124
Table 9 Frequency of Feedback Content Coded.....	134
Table 10 Frequency of Feedback Charge	136
Table 11 Pearson Correlations between Charge Code and FCAT 2.0 Mean Student Achievement Levels.....	138
Table 12 Pearson Correlations between Comment Code and FCAT 2.0 Mean Student Achievement Levels.....	139
Table 13 Research Questions, Variables, Sources of Data, Methods of Analysis, and Results.....	146
Table 14 District percentages of Level 1 and Level 2 Students on the 2013 Reading FCAT 2.0	158
Table 15 District percentages of Level 1 and Level 2 Students on the 2013 Mathematics FCAT 2.0	158

CHAPTER 1

PROBLEM OF PRACTICE

Background of the Study

Accountability is defined as “a policy of holding schools and teachers accountable for students’ academic progress by linking such progress with funding for salaries, maintenance, etc.” (*accountability*, dictionary.com). According to Florida statute 1008.31 (2012),

“The performance accountability system implemented to assess the effectiveness of Florida’s seamless K-20 education delivery system provide answers to the following questions in relation to its mission and goals:

1. What is the public receiving in return for funds it invests in education?
2. How effectively is Florida’s K-20 education system educating its students?
3. How effectively are the major delivery sectors promoting student achievement?
4. How are individual schools and postsecondary education institutions performing their responsibility to educate their students as measured by how students are performing and how much they are learning?”

To hold schools and teachers accountable, student test scores from state and/or district assessments are used in conjunction with instructional practice scores in evaluation models to ensure students are receiving adequate education from effective teachers.

As written in the K-12 education code, statute 1012.34 (2012) states, “Instructional personnel and school administrator performance evaluations must be based upon the performance of students assigned to their classrooms or schools” (Florida Statute 1012.34, 2012). The performance of students is measured using student assessments with “At least 50 percent of a performance evaluation” and “must be based

upon data and indicators of student learning growth assessed annually by statewide assessments or, for subjects and grade levels not measured by statewide assessments, by school district assessments” (Florida Statute 1012.34, 2012). Currently, accountability is relied upon as a strategy to raising academic rigor and measured by student achievement outcomes on high stakes assessments.

Accountability is not a new paradigm. Florida teachers have been held accountable since as early as 1884 when they had to meet the requirements for becoming a teacher through certification and showing “good moral character” (“Assessment and Accountability Briefing Book,” 2007, p. e-8). Through the 1930s and 1940s, results from teacher certification exams were published as “...qualified teachers became permanent employees of their county school system,” and additional levels of certification were given to teachers completing certain requirements in higher learning institutions; the more education and experience a teacher had, the higher the certification level on the license (“Assessment and Accountability Briefing Book,” 2007, p. e-9 & e-10). Beginning in the 1970s, teachers began receiving inservice and training through school districts (“Assessment and Accountability Briefing Book,” 2007). By the 1980s, prospective teacher candidates were required to take the Florida Teacher Certification Exam (FTCE) for certification, and by 1988, teachers needed to pass a subject area exam in addition to a professional exam for certification (“Assessment and Accountability Briefing Book,” 2007). The FTCE added the general knowledge test in 2000 and teachers were required to show additional coursework and a college degree in relation to the subject area they wanted to teach (“Assessment and Accountability Briefing Book,”

2007). A bachelor's degree at the minimum is required for certification, and teachers are responsible for maintaining their certifications through inservice or college credit to recertify. Beginning in 2014, teachers in Florida were required to earn 6 semester hours of inservice or college credit in teaching students with disabilities in order to earn recertification (FIdoe "Florida Educator Certification Renewal Requirements," 2015).

This timeline shows the historical requirements and high expectations for people to become educators. However, in the current culture of education, simply earning the right to call oneself a teacher is not enough to ensure students are indeed receiving a rich education every day in the classroom. Just as an educator monitors student work and progress, teachers' work and progress is also consistently monitored through observation and student achievement outcomes.

Historically, educators have consistently been held to high expectations. The trend in teacher accountability has occurred as rigorous standards continue to evolve, and teachers are held accountable through constant informal and formal observations. The number of required observations needed depends on certain factors; for example, the length of time the teacher has been teaching, the length of employment in a particular county, and the number bargained by teacher associations. The purpose of observations is to give teachers feedback on their pedagogical practices to improve student learning and achievement.

These observations and the addition of student performance scores, lead up to a final evaluation in which the results potentially carry monetary gain. The higher a teacher's evaluation score, the more of a monetary incentive the teacher is eligible to

receive. Conversely, these evaluations also allow for the possibility of releasing teachers. Data from the large urban school district used in this study clearly shows people who are observed, "...can't appeal the content of the observation or the judgment of the observer. Appeal is only possible in the case of procedural error such as going in for a formal observation without a preconference or doing an informal observation and only spending two or three minutes in the room" (Ocps "Teacher Evaluation Frequently Asked Questions," 2011, p. 10). Within the evaluation process, teachers may request more observations if they would like a more reliable score, but they may not grieve the ratings they receive.

This leads to the question of whether evaluations have actually helped teachers by providing necessary feedback that targets how teachers could improve their practices and increase student achievement or if the feedback given is ineffective toward improving student achievement. Research has shown that although teachers are receiving observations on their pedagogical practices, not all feedback given is satisfactory in terms of helping teachers enhance their instruction. According to a study conducted by Rafalski (2015) in an analysis of types of feedback in relation to teacher Value-Added Model (VAM) scores, research showed that the type of feedback whether it be specific actionable feedback or no feedback had little variation on a teacher's VAM score (p.120). While the teachers who received no feedback had lower VAM scores, statistical tests showed there was no significance in the relationship between types of feedback and VAM scores (Rafalski, 2015, p. 120). This may suggest that feedback was not necessary as it did not affect student achievement and therefore the validity of teacher evaluations

may be questioned. VAM is an algorithm that used student historical data to predict what a student's expected growth would be and compared the expected growth with the actual growth, "The teacher's *value-added score* reflects the average amount of learning growth of the teacher's students above or below the expected learning growth of similar students in the state, using the variables accounted for in the model" (Fidoe "Recommendations of the Florida SGIC," n.d., p. 2). VAM scores do not delineate if the student actually shows proficiency on assessments, nor does it tell teachers where student knowledge voids exist. VAM is a reflection on teacher performance through student achievement and as Hattie (2009) states, teachers and school leaders have a direct impact on student performance and achievement. It is important to be able to identify what information can be given to teachers to help them further refine their teaching practices to increase student achievement.

Problem Statement

To date, there has been little research on the content of specific feedback related to teacher evaluation and student achievement data. Recent research has shown that observation feedback provided to teachers through the iObservation protocol did not provide an abundance of specific feedback to teachers that significantly affected student achievement outcomes as measured by the teacher evaluation model in Large Urban School District (Rafalski, 2015). Teacher evaluation systems have been used for accountability purposes, but teachers needed clear, valid, and actionable feedback in order to improve their teaching practices; thereby, improving student achievement.

Purpose Statement

The purpose of this study was to examine observation comments that were coded as specific targeted feedback, and then compare the data to student achievement outcomes. Knowledge gaps that existed between teachings, as noted on the iObservation instrument, and student achievement outcomes were examined. Areas in which most teachers were coded as having received specific targeted feedback and how those areas related to student achievement outcomes were reviewed along with the content of specific targeted feedback to determine whether observer recommendations for improving instruction were aligned with student performance and professional development. The findings of this study will serve to influence policy specifically on teacher evaluation processes and feedback.

Significance of the Study

Findings from this study should give school administrators and district personnel ideas on how to frame feedback for teachers, what specific attention to pay to feedback published to teachers and to the school district, what to focus on during classroom observations, how to follow up and support teacher accountability for specific improvement, and what areas could help improve student achievement specifically for level 1 and level 2 students. Previous research has analyzed the types of feedback given to teachers and found that there is a dearth of actionable feedback that teachers could use to improve their pedagogical practices (Rafalski, 2015). Previous research has also examined the relationships between feedback and Value-Added Model (VAM) scores

(Rafalski, 2015). This research intended to expand upon previous research by examining observation comments coded as specific targeted feedback and analyzing corresponding student achievement data as reported by achievement level outcomes to determine if the specific targeted feedback had any impact on the students' achievement and addressed student knowledge gaps.

Where previous research compared the types of feedback to student achievement through teacher VAM scores, the research did not focus on if students were indeed proficient on assessments. The use of achievement level outcomes was significant because VAM analyzed students' historical assessment data and calculated output as an expected score for individual students to meet. If students did not meet this expectation, the teacher's VAM score was affected. Even if the student passed or scored proficiently on the assessment, having not met the expected score would have affected VAM. The model looked at how a student performed previously on a test, used the variables in an algorithm to predict the next year's growth and then measured the actual difference in growth. If the actual performance exceeded the predicted performance then the teacher received a positive VAM score. If the actual performance fell below the predicted performance then the teacher received a negative score. If the predicted score and the actual score remained the same, then the progress was simply a met expectation (Fldoe "Performance Evaluation," n.d.).

The clinical model was a significant point of interest as the Marzano framework, a growth model, was based on the clinical protocol. As instructional leaders, principals must supervise their teachers through observation and evaluation procedures. This

process helped determine the effectiveness of instructional practice on student achievement. Principals deliver feedback to teacher through formative supervision and summative evaluation. Formative supervision is “The primary method by which principals impact teachers; instructional practice” (Range, Young, & Hvidston, 2013, p. 61). Formative supervision is “proactive” because it gives teachers feedback on their instructional practice and addresses weaknesses to be fixed before teachers receive their summative evaluation, and it functions under the “...assumption that continuous improvement is necessary for teacher growth” (Range, Young, & Hvidston, 2013, p. 61).

Clinical supervision is a commonly used model, which includes a pre-observation conference, observation, and post-observation conference. There is a difference between formative supervision and summative evaluation. Many teachers view supervision and evaluation as the same thing. Formative supervision is an ongoing process for teachers to receive feedback and improve, while evaluation actually rates a teacher’s job performance (Range, Young, & Hvidston, 2013, p. 62).

This study examined information about supervision and evaluation concepts to analyze their implementation in a large urban school district and how student achievement outcomes were affected. The results from this study could help school districts refine their observation processes.

Definition of Terms

The following definitions are presented in five categories to aid in clarification of the terms of this study. The definitions are categorized into areas as they relate to

specific practices or concepts. The categories are: (a) definitions as related to observation, (b) definitions as related to feedback, (c) definitions as related to assessment, (d) definitions as related to professional development, and (e) operational definitions.

Definitions as Related to Assessment

Achievement Levels: “The success a student has achieved with the NGSSS assessed by FCAT 2.0 Reading, Mathematics, and Science is described by achievement levels that range from 1 (lowest) to 5 (highest). Level 3 indicates satisfactory performance” (Fl DOE, “FCAT 2.0 Reading, Mathematics, Science, and Writing Fact Sheet,” 2014, p. 3).

Levels according to the Florida Department of Education (2013):

Level 5 Students at this level demonstrate mastery of the most challenging content of the Next Generation Sunshine State Standards.

Level 4 Students at this level demonstrate an above satisfactory level of success with the challenging content of the Next Generation Sunshine State Standards.

Level 3 Students at this level demonstrate a satisfactory level of success with the challenging content of the Next Generation Sunshine State Standards.

Level 2 Students at this level demonstrate a below satisfactory level of success with the challenging content of the Next Generation Sunshine State Standards.

Level 1 Students at this level demonstrate an inadequate level of success with the challenging content of the Next Generation Sunshine State Standards. (Fl DOE “Understand FCAT 2.0 Reports,” 2013, p. 6).

FCAT 2.0: The Florida Comprehensive Assessment Test. This assessment is given to certain grade levels to test proficiency in an assessed subject area. For the purpose of this study, reading and mathematics will be the only subject areas researched. Students must score a level 3, 4 or 5 to be considered proficient (Fl DOE, “FCAT 2.0 Reading, Mathematics, Science, and Writing Fact Sheet,” 2014). Students in specific

cohorts who never took the FCAT 2.0 assessment had to take the retake examination as well as students who did not initially pass the examination.

Definitions as Related to Observation

Element: The elements in the Marzano protocol are arranged to address specific areas of teaching and learning. Elements as they relate to teaching and learning range from procedural to instructional and consist of teacher-led strategies and student-led strategies. Student-led elements are considered more rigorous. “Research-based strategies are interpreted in the Marzano model as elements. These elements are described through desired effects, and evidence is gathered through teacher and student observed behavior” (Rafalski, 2015, p. 6).

Formal Observation: The clinical supervision model frames formal observations as containing a pre-conference, extended observation, and post-conference. It is in the post-conference where discussions of growth and professional development opportunities are highlighted for teachers to improve instruction (Range, Young, & Hvidston, 2012, p. 63).

Generally, the formal observation is used as the observation for summative evaluation, lasts for an entire class period and provides a rich source of feedback to teachers regarding their instructional practice and professional growth. It includes a pre-conference and a post-conference for reflection with the teacher (RTTT glossary). For the sample district, one formal observation is required annually for a teacher with three or more years of experience in the district. Three years of experience is the point at which under state statute a teacher is no longer considered to be in their developmental period. (Rafalski, 2015, p. 6)

Informal Observation: “Informal observations can be announced or unannounced and typically last from 10 minutes to a full class period. They are used to provide

feedback, track deliberate practice growth, and to collect evidence to inform the annual evaluation process. For the sample district, two informal observations are required annually for a teacher with three or more years of experience” (Rafalski, 2015, pp. 6-7). According to Range, Young, & Hvidston (2012) informal observations are unannounced and, “...occurred at the discretion of principals for initial and continuing contract teachers (p. 66).

Marzano Protocol: As cited by Rafalski (2015) “This protocol consists of 41 key strategies revealed by research for effective teaching presented in a robust, easy-to-understand model of instruction based on *The Art and Science of Teaching* (Marzano, 2007)” (p. 7). Observers are trained to provide observations using this specific framework.

Observer: “Anyone trained and authorized to do informal or formal teacher observations, including rating elements and giving feedback to teachers. This could include principals, assistant principals, instructional coaches, directors, senior administrators or coordinators, as well as other administrators” (Rafalski, 2015, p. 7).

Post-conference: As cited by Range, Young, & Hvidston (2013), the post-conference is the most important part of the clinical supervision model where purposes include “(1) review and reflect upon the data collected during the extended observation, (2) link professional development opportunities to areas of teacher needs or interests, and (3) begin to discuss and preview the next extended observation” (p. 64). The model in large urban school district, Florida only utilizes the pre and post-observation conferences with formal observations.

Pre-conference: A pre-conference as it relates to observation includes the teacher and administrator meeting to discuss what the teacher can expect from the observation process as well as what the observer can expect from the lesson. As cited by Range, Young, & Hvidston (2013), “The purpose of the pre-observation conference is to frame the upcoming extended observation and includes a discussion between principals and teachers concerning the objectives of the lesson, instructional strategies to be used during the lesson, methods by which students will be evaluated and how the lesson will be summarized (p. 63).

Value-Added Model (VAM): As quoted by Rafalski (2015), “In general, value-added models are a class of statistical procedures that use longitudinal test score data, i.e., data collected over a period of time, to measure the change in a student’s performance during a specific period of time’ (Doran & Izumi, 2004, p. 3)” (p. 9). A value-added measure is the metric assigned to specific teachers based on growth in the learning of the students they taught during a specified period of time (Ravitch, 2010) or the difference between the predicted performance and the actual performance represents the value-added by the teacher’s instruction (Florida Department of Education, 2014)” (p. 9).

Definitions as Related to Feedback

Feedback: As cited by Rafalski (2015), feedback is defined as “information about how we are doing in our efforts to reach a goal” (Wiggins, 2012, p. 11), and “feedback needs to provide information specifically relating to the task or process of learning that fills the gap between what is understood and what is aimed to be understood’ (Hattie &

Timperley, 2007, p. 82)” (p. 9). Feedback is “information provided by an agent (e.g., teacher, peer, book, parent, or one’s own experience) about aspects of one’s performance or understanding” and “a consequence of performance” (Hattie, 2009, p. 174).

Feedback Charge: Feedback that possesses qualities that are good, undesirable, or neither good nor undesirable. It can be described as positive, negative, or neutral.

Proficient: When a student scores a level 3 or above on FCAT 2.0. A level 3 is considered a passing score on the FCAT 2.0. The score ranges differ between the assessed subject areas and grade levels (Fldoe, “FCAT 2.0 Reading, Mathematics, Science, and Writing Fact Sheet,” 2014, pp. 3-4).

Specific Targeted Feedback: Rafalski (2015) defined specific targeted feedback as “The observer leaves differentiated and meaningful statements intended to improve the impact of an instructional strategy” (p. 79).

Targeted Feedback: As quoted by Rafalski (2015) “This refers to feedback that is informative, constructive, objective, actionable, and focused on specific classroom strategies and behaviors during a set time interval, (Florida RTTT glossary)” (p. 10).

Definitions as Related to Assessment:

Definitions as Related to Professional Development

Coaching: To take a person from an area that needs work to an area where they are enhancing a skill. “The term coach generally means helping someone move from where he or she is to where he or she needs or wants to be” (Marzano & Simms, 2013, p. 4).

Professional Development: Includes collaboration, trainings, and inservice to provide educators with information and reinforcement of instructional strategies.

Professional development can be a formal training or informal collaborative meeting. It should be “meaningful” and “...tailored to department needs as well as to individual needs” (Gabriel, 2005, p. 113).

The purpose of the professional development system is to increase student achievement, enhance classroom instructional strategies that promote rigor and relevance throughout the curriculum, and prepare students for continuing education and the workforce. The system of professional development must align to the standards adopted by the state and support the framework for standards adopted by the National Staff Development Council. (Florida Statute 1012.98, 2012)

Operational Definitions

Clinical Supervision: A model of supervision that contains a pre-observation, extended observation, and post-observation. (Range, Young, & Hvidston, 2013, p. 62). The model is designed to help teachers become more reflective as opposed to traditional models where supervisors tell the teacher what adjustments to make (Marzano & Simms, 2013, p. 5).

Rubric: “A rubric is a guide for communicating expectations of quality for a task by setting clear criteria and listing specific measures for scoring” (Rafalski, 2015, p. 10)

The rubric for categorizing targeted specific feedback is as follows:

Coaching: The observer gives an explicit coaching tip or refers the teacher to other professionals for support.

Content related: The feedback is related to the specific content or subject-area being taught.

Pedagogy related: The feedback is related to specific teaching practices or strategies.

Procedural: The feedback is related to classroom rules, procedures, and teacher rapport.

Professional development: The observer suggests specific professional development to the teacher to improve instruction.

The rubric for categorizing the charge of feedback is as follows:

Negative: Of or relating to possessing qualities that are undesirable.

Neutral: Of or relating to possessing good or desirable qualities.

Positive: Having no strong good or undesirable qualities.

Social Cognitive Theory: Albert Bandura's theory of self-efficacy. Self-efficacy is people's behavior, which is "...affected by what they believe they are capable of performing" (Owens & Valesky, 2011, p. 296). People are more likely to be successful if they believe they are able to carry out certain tasks. Self-doubt leads people to submit to their own default behaviors.

Theoretical Framework

Prior research explored many theorists and found a predominant trait in reinforcement theory (Rafalski, 2015, p.15). As cited by Rafalski (2015), "This operant of learned behavior is influenced by events and linked to successful performance and self-correction from feedback (Pate, 1977)" (p. 16). For the purpose of this study, Albert Bandura's Social Cognitive Theory was explored in order to better explain why a teacher

may or may not accept and apply feedback from a classroom observation and how that could affect student achievement. A look into supervision and feedback was also necessary as these concepts were related to principles within Bandura's theory. While the concepts of reinforcement theory still apply to this researcher's study, the framework for this study focused on Bandura's Social Cognitive Theory. Social Cognitive Theory states people's behavior is "...affected by what they believe they are capable of performing, which is termed self-efficacy" (Owens & Valesky, 2011, p. 296). If a person has a strong sense of self-efficacy then he or she may "...enthusiastically perform those behaviors," yet if there is no strong sense of self-efficacy then a person may avoid the behavior (Owens & Valesky, 2011, p. 297).

"People are self-organizing, proactive, self-reflecting, and self-regulating, not just reactive organisms shaped and shepherded by environmental events or inner forces" (Bryant & Zillman, 2002, p. 121). It is inherent within people to think about what they have done and how actions can be improved upon. Based on this principle, teachers are constantly reflecting on their practices and making changes based on their own discretions. Supervision models and observation protocols make the reflective process more concrete by delivering written and oral observation feedback to teachers and making suggestions on pedagogical practices. As instructional leaders, principals must supervise their teachers through conducting observations. This process helps determine the effectiveness of instructional practice on student achievement outcomes.

Formative supervision and practices are used by principals to deliver feedback comments to teachers, and even though teachers are given feedback, if the feedback

consists of behaviors or situations that teachers are not able to cope with, then the feedback may not be accepted or applied (Bandura, 1977). According to Bandura (1977) expected outcome is what a person believes will happen if certain behaviors are applied, and what actually happens is the outcome. During an observation if an observer gives specific targeted feedback then an expected outcome would be improved instructional practices, which could lead to increased student achievement. If a person does not think they can cope with the situation, especially "...in the face of obstacles" then the behaviors may not be applied (Bandura, 1977, p. 194). Therefore, if a teacher does not accept feedback because he or she is uncomfortable with the behaviors or believes the obstacles are too great, this outcome could potentially have a negative effect on student achievement.

Taking feedback given by an observer and the ability to apply it to classroom practices is not an easy task given the demands of teaching and accountability. "A number of factors determine whether people will act on what they have learned," such as the environment, incentives, and personal values (Bryant & Zillman, 2002, pp. 145-146). According to Marzano & Simms (2013) "It stands to reason that if a teacher doesn't know what he or she is doing right or wrong, it will be difficult for that teacher to improve his or her knowledge and skill" (p. 10). A difficulty in giving feedback is there is no guarantee teachers will be receptive or open to applying feedback. Changing or correcting an instructional practice may seem like a daunting task and may take time to show positive results. A teacher's self-efficacy will "...determine how much effort

people will expend and how long they will persist” (Bandura, 1977, p. 194). The higher a teacher’s self-efficacy, the more motivated he or she may be willing to try new things.

In linking Social Cognitive Theory to Reinforcement Theory, the longer the teacher persists, he or she may “...gain corrective experiences that reinforce their sense of efficacy, thereby eventually eliminating their defensive behavior;” however, if the teacher cannot cope and does not persist, then the teacher will continue to have fears and low expectations in relation to the behavior (Bandura, 1977, p. 194). Owens & Valesky (2011) state, “Self-efficacy influences the choices we make, how much effort we put into those choices, and how long we persist in our choices if we encounter difficulties...All of this has enormous implications for leaders in the areas of instruction and staff development” (p. 297). Implications for leaders could include how feedback is delivered to teachers as well as what kinds of professional development to recommend in order to make sure teachers are trained in the skills and strategies expected of them.

There are four efficacy expectations related to Social Cognitive Theory. They are: “performance accomplishments” (“enactive attainment”), “vicarious experience”, “verbal persuasions,” and “emotional arousal” (“psychological state”) (Bandura, 1977, p. 195; Owens & Valesky, 2011, p. 297). Performance accomplishments (enactive attainment) is the subject’s own experiences: They may be positive or negative, and the degree to which a person can master a concept without repeated failure will determine how much self-efficacy a person gains (Bandura, 1977; Owens & Valesky, 2011). Vicarious experience is attained through watching others: Modeling and observation are ways that people can learn from others’ success and failures, which may have an impact

on how a person views his or her own ability to perform the same tasks (Owens & Valesky, 2011, p. 297). Verbal persuasion is directly linked to feedback. This is information that a person received from others, whether it is positive or negative and is “...more effective when the person giving feedback is viewed as skilled in performing the same task, and when the feedback is presented positively” (Owens & Valesky, 2011, p. 297).

For the purpose of this study, positive, negative, and neutral feedback were defined and explored. The question begs if teachers will more openly accept corrective feedback if it is presented positively. Emotional arousal (physiological state) determines how stress and anxiety affect a person’s ability to perform a task (Owens & Valesky, 2011, p. 297). Observations can be stressful for teachers who are worried about the relationship of student outcomes to their performance. With the question of accountability and performance pay, teachers may have other worries about observations that hinder their ability to process and implement feedback successfully. According to Owens & Valesky (2011) “All of this becomes important to organizational leaders who need to be aware of the status of employee self-efficacy. Employees who perceive themselves as highly efficacious will be more likely to put forth effort that is sustainable to produce successful outcomes, and of course, the reverse is true” (p. 298).

Bandura’s theory states stress can hinder performance, “Because high arousal usually debilitates performance, individuals are more likely to expect success when they are not beset by aversive arousal than if they are tense and viscerally agitated” (Bandura, 1977, p. 198). Stress may also deter a teacher from accepting feedback. Because of the

highly evaluative nature of observations, teachers must not feel they are in a threatening or evaluative environment to be receptive to feedback. Marzano & Simms (2013) state, "...effective communication can only take place in a nonthreatening environment..." (p. 9). Anxiety affects performance, which in turn may affect a teacher's efficacy.

A performance accomplishment (enactive attainment) is one type of influence under Social Cognitive Theory that reduces defensive behaviors as they "...promote behavioral accomplishments but also extinguish fear arousal, thus authenticating self-efficacy" (Bandura, 1977, p. 195). Self-efficacy rises as successes rise and fall as failures are repeated; however, as someone is able to see how they are able to master an obstacle, failures can be overcome (Bandura, 1977, p. 195). Giving a teacher feedback may elicit feelings of failure if the recommendation or strategy does not work immediately. As teachers try new strategies and do not feel the interventions are working, then teachers may feel like they are not able to use the feedback or suggestions given. Observations are essentially performance assessments of teachers. A goal of observations is for teachers to become more reflective of their practice and therefore address knowledge gaps to enhance student achievement outcomes. Often people do not think about how their personal mastery may be affected and "...moreover, expectations are usually assessed globally only at a single point in a change process as though they represent a static, unidimensional factor" (Bandura, 1977, p. 194). During an observation an administrator may only watch a snapshot of a lesson and therefore not see the full extent of a teacher's performance mastery or growth. Conversely, a teacher may compliantly follow an instructional strategy protocol solely for the purpose of observation. This

illustrates Bandura's point about self-efficacy and personal mastery. As a teacher practices a skill or strategy for the purpose of personal mastery then observation becomes less of a show of compliance and more of an opportunity to showcase strengths, which in turn makes teachers viable models for other struggling teachers.

Anything that affects performance is going to affect efficacy. Bandura's (1977) Social Cognitive Theory claims performance is the best form of boosting efficacy because they are authentic experiences. Teachers may be more likely to apply feedback and try instructional strategies if they have had their own successes. While verbal persuasion and vicarious experiences are helpful in giving teachers the confidence they may need to try something different or new, they are not as strong in raising self-efficacy than if a teacher is able to experience the results him or herself. Therefore, it is important for teachers to be open to feedback and proactive enough to seek ways to raise their efficacy. Whether it be trial, modeling, or receiving feedback, applying the strategy or behavior is necessary to even begin raising self-efficacy. Social cognitive theory creates somewhat of a cycle. Teachers must be receptive and accepting to feedback, apply feedback, and have enough success to make behavior changes, which could positively affect student outcomes. However, if teachers are defensive and do not have the coping mechanisms necessary to make behavior changes, students could be negatively impacted as teacher self-efficacy falls (Bandura, 1977).

Research Questions

The following are research questions as they related to this study:

1. What is the frequency of level 1 and 2 students in relationship to teachers who received specific targeted feedback?

2. What relationships, if any, exist between the specific targeted feedback as measured by the elements scored during a school year and student achievement outcomes as measured by FCAT 2.0?

H₀₁. There are no significant relationships between the type of element scored during observations during a school year and student achievement outcomes as measured by FCAT 2.0.

3. What is the frequency by category of feedback, defined as content related feedback, pedagogy related feedback, procedural related feedback, coaching related feedback, or professional development related feedback provided by observers to teachers during classroom observations?

4. What relationships, if any, exist between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT2.0?

H₀₂. There are no significant relationships between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0.

5. What relationships, if any, exist between the categories of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0?

H₀₃. There are no significant relationships between the categories of specific targeted feedback and student achievement outcomes as measured by FCAT 2.0.

Limitations

The following were limitations that could have potentially affected the results in this study:

1) There were variables, other than the teacher, that were outside the control of the researcher that could affect student achievement. While Hattie (2009) says teachers have a 0.44 effect size on student achievement (p.115), other factors such as socioeconomic status, learning disability, and English language proficiency could affect how a student performs on an assessment.

2) There were also factors that may have affected a teacher's instructional practice score during an observation such as the potential of observer bias and rater reliability.

3) Limitations of student data in relationship to teacher practices included students who may have been in a teacher's class for only a limited amount of time.

4) Data did not indicate how long students were with specific teachers. Since the data were taken from a previous study, limitations from that study also applied to this study.

5) The researcher could not control for the maximum number of observations each teacher received, and only electronic feedback was used (Rafalski, 2015, p. 19).

6) Other limitations related to the data used in this study were the grade levels associated with the FCAT 2.0. The mathematics assessment only tested grades 3-8, the

reading assessment tested grades 3-10, and the reading retake assessment only tested students in grades 11 or 12 who did not take or pass the assessment during the initial school administration.

7) While the researcher did read and code feedback comments to determine what categories of specific targeted feedback existed, the accuracy and quality of the feedback were not focal points for this study.

Delimitations

The study was delimited by the following:

1) A large urban school district in Central Florida was chosen for the location of the study. This school district employed approximately 14,000 teachers and educated approximately 192,000 students.

2) The researcher reviewed a sample from one year of observation data that was previously coded through Rafalski's (2015) study, which was the origin of this researcher's study. The original sample size consisted of 2,718 teachers based on criteria of only using teachers with at least three years of experience, National Board Certification, and was limited to elementary, middle, and high school teachers. (Rafalski, 2014, pp.17-18).

3) Rafalski's study (2015) used a rubric to determine types of feedback given to teachers during observations and found of the 2,718 teachers, only 91 teachers fell into the category of having received predominantly specific targeted feedback. The current

study examined the sample of 91 teachers who were rated as having received predominantly specific targeted feedback.

4) This study utilized one year of reading and mathematics FCAT 2.0 data from the 2013-2014 school year, which included students who took FCAT 2.0 as a retake assessment.

5) Data did not identify if teachers were elementary, middle, or high school teachers.

6) Achievement levels were used instead of developmental scale scores since scale scores differ among grade levels whereas all achievement levels measure proficiency on a 1-5 scale.

7) This study did not delineate the subject-area the teachers taught.

8) All teachers who received specific targeted feedback were included in this study as well all students connected to these teachers who took the FCAT 2.0.

Assumptions

This study had several assumptions.

1) This study assumed that the observers who were evaluating teachers were trained in the Marzano protocol and followed the clinical model of supervision (formal observations) in which observations included a pre and post conference.

2) This study also assumed that principals were aware or somehow made aware of professional development opportunities for teachers and school-based coaches were utilized to help teachers improve their pedagogical practices.

3) If an observer did give specific targeted feedback, it was assumed that teachers had access to their feedback prior to the next observation as required by the Orange Classroom Teachers Association (CTA) contract (2013-14, p. 44).

4) It was assumed trained observers conducted all observations, and although not delineated in this study, all formal observations were conducted by a trained administrator per CTA contractual obligations (2013-14, p. 44).

5) It was assumed that principals and observers had a working knowledge of instructional best practices and were aware of student data as it related to the school.

6) Assumptions about students included any FCAT 2.0 assessment score not coded, as a retake assessment was the students' initial attempt on the grade level test.

7) For those students who needed to retake the FCAT 2.0 Reading Assessment, the assumption was they were in the cohorts of students who needed FCAT 2.0 as a graduation requirement and did not initially pass, or they had never taken the assessment before.

8) For students who did not pass the FCAT 2.0 Reading assessment, the assumption was they were previously level 1 and 2 learners as measured by FCAT 2.0 achievement levels.

9) Assumptions about the observation feedback were that the comments were easily understandable, that the Marzano elements were correctly coded, and the feedback offered valid and actionable instructional strategies.

10) Finally, it was assumed the comments coded as specific targeted feedback according to Rafalski (2015) were indeed specific targeted feedback. The scope of this research did not consist of rechecking comments coded from Rafalski's (2015) study.

Organization of the Study

This study was organized into five chapters. Chapter 1 was an overview of the study and included the problem statement, the purpose of the study and its significance to the body of research, the conceptual framework which guided the research, limitations which may affect how results are interpreted, delimitations which were the boundaries set by the researcher for the purpose of the study, and assumptions which may influence recommendations for further research. Chapter 2 provided a review of relevant literature and research. Chapter 3 reported the methodology of the research, and chapter 4 provided the findings and analysis of data from the research. Chapter 5 was the final chapter in this study, which summarized the data and discussed the implications for educational practice as well as recommendations for future research.

Methodology

This study was a mixed-methods analysis, which examined a portion of data previously coded by another researcher, Rafalski (2015). Quantitative and qualitative procedures were used to answer 5 research questions. Descriptive statistics were calculated to identify frequencies within the data for the purpose of answering specific research questions. Categories of specific targeted feedback were created using a sub-rubric in Rafalski's (2015) study. Rubrics were created to determine the categories of

specific targeted feedback (content related, pedagogy related, procedural related, coaching related, or professional development), and to determine the charge of feedback (positive, negative, neutral). The feedback data were rated on with the rubrics and coded by category and charge of feedback given to teachers during observations. The data were compared to corresponding student achievement level data to identify specific areas where there were student and teacher knowledge gaps. Pearson's r was calculated to examine statistically significant relationships between observation data and student achievement level outcomes as measured by FCAT 2.0. The purpose of this analysis was to determine which categories and charges of specific targeted feedback were most beneficial to student achievement outcomes.

Procedures

This study was executed in a multi-step process in order to accurately answer the 5 posed research questions. First, a sample of data was collected from the 2013-2014 school year. These data included 91 teachers who were previously coded as having received specific targeted feedback in observations (Rafalski, 2015) and their corresponding students' achievement level outcome data. The researcher used descriptive statistics to analyze the frequencies of level 1 and level 2 students for each FCAT 2.0 assessment (reading, reading retake, and mathematics). Next, the researcher examined the observation protocol map and the frequencies the observed and recorded elements to determine in which areas observers provided feedback.

A sub-rubric from Rafalski's (2015) study was used to determine specific categories of feedback. A rubric was created to examine the raw observation comments. The comments were coded and sorted into categories of specific targeted feedback (content related, pedagogy related, procedural related, coaching related, or professional development related). Another rubric was created to determine the charge (positive, negative, neutral) of the observation comments. The comments were coded and sorted into the specific charge category of feedback.

. Descriptive statistics were used to examine the frequencies of categories of specific targeted feedback and charges of specific targeted feedback. The researcher then calculated the averages of feedback categories and feedback charges for each teacher to create quantitative measures used for calculating Pearson's r . Pearson's r was then calculated to analyze if statistically significant relationships existed between the categories of specific targeted feedback and student achievement level outcomes as measured by FCAT 2.0. Pearson's r was calculated to also examine if statistically significant relationships existed between the charges of specific targeted feedback and student achievement level outcomes as measured by FCAT 2.0.

Population and Sample

The population of this study was a sample of 91 teachers and their corresponding students in a large urban school district in Central Florida. Data were used from a previous study (Rafalski, 2015), which analyzed different types of feedback and the significance to teacher VAM scores. The purposive sample was analyzed to further

examine the content of the specific targeted feedback and how it may have affected student achievement outcomes on high stakes assessments.

Generalizability of results could encompass all schools in the large urban school district in Central Florida and other school districts in Florida using the same Marzano protocol and student assessments.

Data Collection

The purposive data sample collected met specific criteria as it related to specific targeted feedback (Rafalski, 2015). A previous study examined a sample of 2,718 teachers and narrowed the sample to 91 teachers who received specific targeted feedback (Rafalski, 2015, p. 18). Rafalski (2015) defined specific targeted feedback as, “The observer leaves differentiated and meaningful statements intended to improve the impact of an instructional strategy” (p. 11). Teachers that were coded as having received predominantly specific targeted feedback were further examined to research student achievement outcomes.

Historical student assessment data for the 2013-2014 school year was collected to examine the significance of relationships between categories and charges of specific targeted feedback to student achievement level outcomes. The student data were the FCAT 2.0 Reading, Reading Retake, and Mathematics assessment scores that were connected to the 91 teachers in the sample. The source of these data came from the district database in the Central Florida large urban school district.

Data Analysis

Research question 1 sought to determine the frequency of level 1 and 2 students as instructed by teachers who received predominantly specific targeted feedback. Student achievement outcome data for students who received a level 1 or 2 on the FCAT 2.0 Reading, Reading Retake, and Mathematics assessments were analyzed. Descriptive statistics and frequencies were used to determine how many students received a level 1 or 2 on FCAT 2.0 even though the teachers received specific targeted feedback.

Research question 2 sought to examine what relationships, if any, existed between the specific targeted feedback as measured by elements scored during a school year and student achievement outcomes as measured by FCAT 2.0 Reading, Reading Retake, and Mathematics assessments. Frequencies were calculated to examine how many of each Marzano element was coded to identify gaps. Pearson's r was calculated to examine the significance of the relationships between which elements were scored in an observation and student achievement outcomes for teachers who received specific targeted feedback.

Research question 3 sought to determine the frequency by category of feedback, defined as content related feedback, pedagogy related feedback, procedural feedback, coaching feedback, or professional development feedback provided by observers to teachers during classroom observations. Data were coded using a rubric (appendix d) and analyzed using descriptive statistics and frequencies to determine the occurrence of each type of feedback.

Research question 4 sought to examine what relationships, if any, existed between the frequencies of positive, negative, or neutral feedback and student achievement

outcomes as measured by FCAT 2.0. Data were coded using a rubric (appendix e) and descriptive statistics were run to calculate frequencies. Correlations using Pearson's r were analyzed to examine the significance of relationships between the charge of the feedback and student achievement outcomes.

Research question 5 sought to examine what relationships, if any, existed between the categories of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0. Pearson's r was used to examine the significance of the relationships between the categories of the feedback and student achievement outcomes.

CHAPTER 2 REVIEW OF LITERATURE

Introduction

This chapter explains the purpose for conducting research on the relationships between teacher observation feedback and student assessment scores as measured by FCAT 2.0. Accountability and evaluation are tools widely used to monitor the effectiveness of teachers in the classroom. Effectiveness of teachers is measured by multiple factors with the most emphasis placed on student achievement. While specific formulas and value-added models are used to calculate student performance for evaluation purposes, many of these models have failed to include outside variables that may affect student outcomes. A student may pass an assessment, but not meet an expected result as calculated through an algorithm and negatively affect a teacher's evaluation score.

This research sought to examine assessment, evaluation, and supervision to determine what kind of quality feedback was needed to support student achievement. The purpose of evaluation was to determine teacher effectiveness in the classroom. These results could be used to make personnel decisions and may affect pay. Supervision, on the other hand, is a tool to build capacity in teachers, thereby improving student achievement through feedback, coaching, and professional development. The Marzano instructional framework as used is a tool that combined supervision and evaluation, a difficult task.

Research supports the necessity of feedback that is specific, timely, and relevant to teacher needs in order to be effective. Supervision is a method in which coaching and feedback could be delivered to teachers. Within feedback, supervisors should be providing teachers with coaching advice, professional development opportunities, and structuring feedback in a way that promotes teacher efficacy and motivation.

This research began by examining a study previously completed on the types of feedback given to teachers in a large urban school district. Research showed that only a small percentage (3.35%) of feedback was predominantly specific and targeted to help the teacher (Rafalski, 2015, p. 119). This research aimed to go deeper by examining the percentage of teachers who received specific targeted feedback and analyzing the content of the feedback. The research then determined if there were statistically significant relationships between the content of specific targeted feedback and student achievement outcomes. The process of determining which literature to focus on began by reviewing the sample rubric of key words found in specific targeted feedback of Rafalski's (2015) study. The rubric can be found in appendix c of this researcher's study.

Multiple searches were performed through the UCF library from databases such as ERIC, Education Full Text, and Professional Development Collection Education (EBSCOhost). The following topics for this literature review as they related to the research questions were: (a) accountability and assessment, (b) evaluation, and (c) supervision. The topics had specific subtopics and foci to further narrow the scope of the research.

Accountability and assessment covered a brief history of assessment with the focus narrowed to FCAT 2.0. FCAT 2.0 was the focus assessment of this study as those were the test results being analyzed in relationship to teacher observation feedback. Evaluation encompassed the elements of evaluation models in Florida and the large urban school district of this study, the Marzano instructional framework, and value added models. The Marzano instructional framework was the evaluation tool used in the large urban school district for observation and evaluation purposes. A value-added model was also used to calculate a teacher's student growth score which was used as part of his or her evaluations. Supervision included subtopics on feedback, coaching, and professional development.

Research was found through books, articles, legislation, and web documents from state sites. The researcher reviewed the references sections of articles and books to further research the topics and subtopics, and looked for recent research related to the topics. Through this method, the researcher discovered that the research began repeating itself through the authors, experts, and the content of the research. While there was literature of research on these topics to support this dissertation, there was little empirical research; therefore, further research is needed to determine the effectiveness of specific types of quality feedback given to teachers through supervision and evaluation protocols. These topics were discussed in separate sections; however, there were many overlapping elements and connections made throughout the literature review.

Accountability and Assessment

Accountability is a response to the demand for "...increased efficiency in the operation of the public schools" (Brimley, Jr., Verstegen, & Garfield, 2016 p. 33). Some concerns with accountability are that lawmakers may not take into consideration the consequences of accountability laws on teachers and students (Brimley, Jr., Verstegen, & Garfield, 2016). Some of these "questionable notions" include students being tested and the results being used to determine the effectiveness of teachers and schools (Brimley, Jr., Verstegen, & Garfield, 2016).

Accountability and assessment have traditionally been woven into the fabric of education. The assessment data examined for this research is FCAT 2.0. The following is a timeline regarding the beginnings of assessment and accountability in Florida and implications for schools, teachers, and students. The timeline shows how assessment has changed over the last 50 years and illustrates why FCAT 2.0 was the data chosen to be analyzed.

As early as 1963, lack of information about student achievement was a concern, and around this time Florida began working toward assessment of students (Fl DOE "Assessment and Accountability Briefing Book," 2007). In 1968, the state made a requirement to improve the effectiveness of education in Florida, and one year later, the Research and Development Program was established, which created assessment items and "contributed to Florida's accountability efforts" (Fl DOE "Assessment and Accountability Briefing Book," 2007, p. e-1). At this time, Florida began creating new assessments for reading, writing, mathematics, and science.

During the 1970s, students in elementary through high school were assessed in multiple subject areas as well as given the first graduation examination or high stakes tests, while 1978 saw the first legal challenges to state assessments (Fldoe “Assessment and Accountability Briefing Book,” 2007). Concerns included high-stakes graduation requirements and test validation. Assessments given in the 1980s determined whether a student would receive a high school diploma, and through a response to a court case it was decided Florida could indeed, “deny diplomas” to students who did not pass the assessment (Fldoe “Assessment and Accountability Briefing Book,” 2007, p. e-5). The significance of denying diplomas to students from this court case has implications in Florida schools today. If a student does not pass the state assessment, he or she may not be awarded a diploma or graduate from high school. Students, who did not pass the FCAT 2.0, could not graduate with a diploma unless a concordant score on a national assessment (such as ACT, SAT) was achieved or a waiver as related to an individual education plan (IEP) was granted.

School reforms of the 1960s and 1970s generated the writing of a report called *A Nation at Risk* (Ravitch, 2010). *A Nation at Risk* criticized education and received notorious attention. During this time, teachers were blamed for low student performance on assessments. Among the problems that *A Nation at Risk* addressed was “teacher preparation,” and a recommendation in the report called for “higher standards for entry into the teaching profession” (Ravitch, 2010, p. 25). These teachers should meet high standards and prove their “competence in an academic discipline” (Ravitch, 2010, p. 27). In return for showing their competence, it was argued, teachers’ salaries and tenure

should be raised but “performance-based” and tied to “peer review” (Ravitch, 2010, p. 29). Performance-based pay for teachers is generally based upon how well a teacher’s students perform on mandated or high-stakes assessments. Performance can be based on pass rates or growth and determine if schools receive extra money that can be used for teacher bonuses.

In 1984, The State Board of Education raised student performance standards to “encourage students and teachers to attain higher achievement” (Fldoe “Assessment and Accountability Briefing Book,” 2007, p. e-6), and later achievement level data from assessments were reported. In 1991, the Florida Commission of Education Reform and Accountability called for “higher levels of achievement,” “more accountability for schools,” sought to “reward higher performing schools,” and provide “assistance to unsuccessful schools” (Fldoe “Assessment and Accountability Briefing Book,” 2007, p. 1). As expectations increased, accountability for schools meant assessment scores would be used to determine if schools needed state intervention and could affect the funding of a school.

The Florida Comprehensive Assessment Design (FCAD) was introduced in 1995 in order to “raise expectations for students” and assist them in career readiness (Fldoe “Assessment and Accountability Briefing Book,” 2007, p. 2). College and career readiness has been a focus up through recent legislation in order to better prepare students for post-secondary education and jobs. This trend is evidenced in the continuous changes in educational standards, such as the adoption of Common Core. In 1996, Florida adopted the Sunshine State Standards (SSS) and authorized the Florida Comprehensive

Assessment Test (FCAT), which was "...field tested in grades 4, 5, 8, and 10" for reading and mathematics in 1997 (Fldoe "Assessment and Accountability Briefing Book," 2007, p. 2). The first actual administration of FCAT was given in 1998, but the results were not reported for accountability until 1999 when the FCAT was used to first assign school grades (Fldoe "Assessment and Accountability Briefing Book," 2007, p. 3). In 2000, the first norm referenced FCAT was given to see how Florida students performed compared with other students in the nation, and in 2001, the Board of Education determined the FCAT cut scores that students needed for graduation (Fldoe "Assessment and Accountability Briefing Book," 2007). During this year The No Child Left Behind Act was passed, which required annual assessment for students in grades 3 through 8 and one time in high school (Fldoe "Assessment and Accountability Briefing Book," 2007). The significance of determining cut scores has implications in education as the cut score becomes the standard to reach so students can pass high stakes assessments. The focus on assessments and cut scores may determine how a teacher teaches students and where students are placed in classes for remediation.

In January of 2002, the Bush Administration signed the No Child Left Behind Act (NCLB) into legislation. No Child Left Behind (NCLB) was quoted as being "the most sweeping reform of the Elementary and Secondary Education Act" and touts "accountability for results" (Usdoe "Fact Sheet on Provisions of NCLB," 2003). This act required the creation of assessments in each state, that each state define its own goals, and determine their own proficiency levels for passing the test. "Student progress and achievement will be measured according to tests," and the data from these tests were

“available in annual report cards” (Usdoe “Fact Sheet on Provisions of NCLB,” 2003). These report cards gave information on the quality of schools and teachers and assigned a letter grade to each school.

NCLB made standardized testing the “primary measure of school quality” (Ravitch, 2010, p. 15). Reading and mathematics assessments were used to measure student proficiency. The validity and reliability of such tests, however, were not given much thought (Ravitch, 2010). As new high stakes assessments have been introduced into education, validity and reliability are variables that stakeholders have consistently monitored because of the implications of such assessments on students and communities.

Annual growth scores (learning gains) were first reported in 2002, which allowed students, parents, and teachers to see how a student progressed over the course of a year. The graduating class of 2003 was required to pass FCAT in order to graduate high school (Fldoe “Assessment and Accountability Briefing Book,” 2007, p. 5). If students did not achieve the score needed on FCAT to pass, they were not able to receive their diplomas, much like the practice on the 1980s.

Since NCLB was enacted, Florida has undergone many changes. First, Florida had Sunshine State Standards (SSS) and tested with the Florida Comprehensive Assessment Test (FCAT). Florida then adopted the Next Generation Sunshine State Standards (NGSSS) and began testing with FCAT 2.0. Scores were organized into proficiency scales so they could be compared across grade levels: students needed to achieve a level 3, 4, or 5 to be promoted to the next grade level or graduate (Fldoe “Understanding FCAT 2.0,” 2013). FCAT 2.0 scores were used in this research to study

the relationships between student achievement outcomes and feedback given to teachers, as this was the assessment used during the 2013-2014 school year. While assessments have changed since the beginning of the original study, the purpose was not to focus specifically on the test, but on the relationship between specific targeted feedback and student achievement outcomes. Since the expiration of FCAT 2.0, which occurred after the 2013-2014 assessment, new standards have been adopted in Florida, which were based on the Common Core Standards. With these new standards have come new assessments, concerns with validity and reliability, and concerns with the state's accountability system.

Testing can be a helpful tool to measure student growth and monitor progress. The results can be used to tell teachers in which areas students may need improvement. Today, technology is available to tie teachers to specific students for accountability purposes (Ravitch, 2010). Ravitch (2010) describes accountability as either "positive" or "punitive" (p. 163). "Positive accountability" is described as scores being used to trigger help for the school, and "punitive accountability" is described as giving a reason to punish or fire personnel because of low scores (Ravitch, 2010, p.163). This will be further discussed in the evaluation and supervision sections of this literature review.

Assessment and accountability have been part of the educational system for decades. The movement towards high-stakes and rigor is continuously increasing. The importance of preparing students for college and careers as well as maintaining competition with other nations are driving forces of holding schools and teachers accountable for student growth and achievement. While legislation showcased

accountability is an innovative key to solving the country's education problems, history has shown that it is nothing new. While accountability is not a new paradigm, federal and state legislation has focused on making schools better by holding teachers accountable through student testing. To hold teachers and schools accountable, procedures have been developed to quantify teaching into scores for evaluation purposes.

Evaluation

Teacher evaluation is a tool used for educational accountability; however, there are problems with evaluation as there is not a clear increase in improving teaching as a result (Tutyens & Devos, 2013). Problems identified with evaluation included teachers being given good ratings when they may not in fact deserve them, teachers lacking meaningful, individualized, and quality feedback, teachers not guided towards relevant professional development, and school leadership's lack of quality time they need on evaluation because of other duties that need attention (Tutyens & Devos, 1998). Legislation has laid the framework for testing and accountability. Evaluation is also a part of this legislation, as teachers are being held accountable for student performance through assessment and measured through evaluation systems.

Race to the Top was education reform enacted under the Obama Administration. According to reasons outlined in RTTT, the Obama Administration believed America needed to do a better job of educating students. To do this, schools needed to set higher standards, create more challenging assessments, and recruit better teachers. July 24, 2009 President Barack Obama said, "...if you set and enforce rigorous and challenging

standards and assessments; if you put outstanding teachers at the front of the classroom...your state can win a Race to the Top grant” (“Fact Sheet: The Race to the Top,” 2009). Under this reform, schools needed to help students prepare for college and career readiness. States could apply for a grant that would help schools achieve these goals. In order to be eligible for the money, states had to link data on student achievement or “growth to teachers and principals for evaluation purposes” (“Fact Sheet: The Race to the Top,” 2009). Race to the Top also encouraged data-driven instruction. In response to Race to the Top, states, in addition to Florida, including California, Indiana, Wisconsin, New York, Delaware, Texas, Connecticut, and many others have either considered or created evaluation systems that linked student performance on assessments with compensation programs for effective teachers (“Fact Sheet: The Race to the Top,” 2009).

While the U.S. Department of Education held high expectations for teachers and required student growth to be a “significant” factor of evaluation, it did not believe student achievement outcomes should be the sole means of evaluation. The Department believed that evaluations for teachers should examine “multiple measures” but the additional measures should be left up “to educators and leaders in LEAs and/or states who are close to the classroom and who can best determine which metrics work in their environments” (Usdoe “Race to the Top Program,” 2010, p. 21). While specific accountability measures was one of the requirements to receive federal RTTT dollars, the power of choosing those measures was left up to the states and local school districts. Florida uses a combination of an observation protocol and student growth from

assessments to measure teacher effectiveness. This effectiveness has implications on job retention and bonus opportunities.

Student achievement and student growth, while seemingly synonymous are very different. “Student achievement is the foundational definition and is structured to give LEAs and States the flexibility to develop measures for all subjects and all grades” (Usdoe “Race to the Top Program,” 2010, p. 22). Student achievement is what a student achieved, a score or raw data that has yet to be defined, it is what student growth is based upon; “Student growth is defined generally as the change in student achievement between two or more points in time” (Usdoe “Race to the Top Program,” 2010, p. 22). Student growth is how much students progressed from year to year, indicating how much they actually learned.

A goal of Race to the Top, like NCLB, was to recruit high-quality teachers in high needs areas. If the Department of Education required teachers to show significant student achievement in order to be considered an effective teacher, then that could potentially discourage teachers from going to high needs schools; therefore, “*growth*” was the “...significant factor in determining effectiveness” (Usdoe “Race to the Top Program,” 2010, p. 21).

While Race to the Top was a federal measure to reform education, each state was able to tailor how Race to the Top would be implemented. Florida’s Race to the Top included evaluating teacher accountability while holding students to rigorous standards for college and career readiness, but with additional requirements. Senate bill (SB) 736, or the Student Success Act (SSA), was a more specific set of criteria for personnel

evaluation purposes. Where there were discrepancies between Race to the Top and the SSA, the SSA would “supersede” as it was a state act (American Institute for Research, 2011, p.7).

Under the SSA there were a few differences in comparison to requirements in Race to the Top. Both Race to the Top and the Student Success Act required states to account for 50% of student growth on teacher evaluations. SSA specified the number of years of previous data (3) that would be used to measure achievement as well as what to do if that data did not exist. SSA also specified the factors that could and could not be used in the achievement formula. Under SSA, the Commissioner was allowed to weight assessments as well as create new formulas as new assessments, such as the End of Course exams (EOCs), were implemented. SSA established performance levels for what standards needed to be met in order to achieve an “effective” or “highly effective” status. The way teachers are evaluated was determined by a group of stakeholders who researched different models for teacher evaluation purposes. .

The Student Growth Implementation Committee (SGIC) members were a diverse group of stakeholders, leaders, and community members who represented “diversity in culture, community, and region” (Fldoe “American Institute for Research,” 2011, p. 5). The goal of the committee was to find a “valid and reliable measure” of student growth for evaluations and to make sure the measures were understandable by personnel and the community (Fldoe “American Institute for Research,” 2011, p. 17). They were appointed by the Commissioner of Education to work on choosing and recommending a model to evaluate Florida teachers statewide. The members looked at eight different models that

could continue to be analyzed based on data and feedback (Fildoe “American Institute for Research,” 2011).

Florida state statutes were the guiding force behind the evaluation policies specific to Florida. The statutes were echoes of what the Student Success Act required. Statute 1012.34 outlined personnel evaluation procedures and criteria. This statute stated the district superintendent would establish evaluation procedures for the purpose of improving personnel (instructional) quality (§1012.34, 2012). Each school district in Florida was able to choose the evaluation system best suited for its teaching and learning community. This is evidenced by the language in the statute and the existence of multiple evaluation platforms in the state of Florida.

According to the 2012 statute (§1012.34), the evaluation systems must support effective instruction and student learning growth, and must provide for continuous growth of instructional skills. The systems must look at different data sources and include different performance levels: highly effective, effective, needs improvement, or unsatisfactory (§1012.34). State statute also required those who were evaluating others, to become trained evaluators. Those who evaluate using the Marzano protocol in Large Urban School District must take a four-day training and pass qualifying assessments to use the model. Evaluators can be any supervisor in a school who was trained on the evaluation protocol and given permission by the principal. These evaluators can be deans, instructional coaches, or even department chairs. An evaluator does not have to be in an administrative role to supervise; however, stipulations may have been bargained between the school district and teacher’s union. Consistency and effectiveness in use of

the observation tool must be monitored; therefore, inter-rater reliability may have been a strategy used in schools by allowing assessing supervisors and administrators to observe teachers throughout the school, not only those assigned through content areas. For example, an assistant principal who usually evaluates Language Arts teachers may evaluate Mathematics and Science teachers in an effort to test the reliability of the evaluation model. If teachers received similar evaluations from different administrators, then that shows consistency in what was occurring in the classroom.

Classroom teachers' evaluations must be based on students taught by them or assigned to their schools. Evaluations may include other evaluative measures; they do not have to be based on student performance alone. In Large Urban School District, these other measures included instructional practices, planning, and professionalism. The numbers of evaluations received were based on the teacher's experience in the district. Experience in the district was key as a veteran teacher may come in from another district and be observed more than a veteran teacher within the district. Not all districts in Florida use the same evaluation model. In 2013, 50% of the evaluation was based on student performance and also measured specific instructional practices (§1012.34). There was also a process established to "monitor school district implementation of evaluation systems," and rules related to student growth standards that would result in the employee earning either an "unsatisfactory," "effective," or "highly effective" rating (§1012.34, 2013).

Evaluation is used for accountability purposes. Student outcomes and other variables were used to measure teacher effectiveness. While federal and state legislation

may require evaluation systems, LEAs had some choice in which system was chosen. In Florida a combination of a specific district chosen observation protocol and a state required value-added component was chosen for the evaluation model requirement.

VAM

A value-added model (VAM) is a growth model that measures changes in student performance from year to year or “test to test” (Fldoe “American Institute for Research,” 2011, p. 30). VAM is tried to capture the learning of the student by controlling for specific variables so that one teacher or school does not have an advantage over the other simply because of the students they have. Value-added models compare student performance with the expected outcome calculated based on previous data. For VAM, a value-added estimate is the “teacher contribution to student learning;” (Fldoe “American Institute for Research,” 2011, p. 29).

Value-added estimates assume that “deviations from expectations among students taught by a teacher are caused by the teacher” (Fldoe “American Institute for Research,” 2011, p. 42). This means student performance is a result of what occurs in the classroom with the teacher. Teacher effects are only estimated, not exactly measured; therefore, it is important to consider how much damage can be done by “mislabeling a teacher” (Fldoe “American Institute for Research,” 2011, p. 61). Labeling a teacher as unsatisfactory has implications with performance pay and job security. A goal of the SGIC was to find a model that produced the lowest amount of risk of misclassification

and carefully accounted for teacher and student contributions to achievement and performance (Fidoe “American Institute for Research,” 2011).

Historically, to track student growth, school officials began using “growth models” (Ravitch, 2010, p. 179). Ravitch, (2010) described the “value-added assessment” (VAA) as a model developed by a statistician named William Sanders who aimed to calculate different factors, including the teacher, contributing to student achievement. By doing this, test scores could be tracked from year to year and target the specific years and teachers whose scores made gains or loses. The problem with value-added assessment was it was purely data driven and does not take into account the curriculum, instruction, or student experiences (Ravitch, 2010, pp. 179-180). As quoted by Ravitch (2010), Dale Ballou said VAA is “dangerous” when used to make personnel decisions because a student’s emotions can affect them on test day, something trivial like the weather could potentially affect students, and statistical “measurement error” could affect scores (p. 184). The results of a VAM may also depend on the variables used in the algorithm.

Sanders believed teaching could be “...quantified by taking students’ test scores, plugging the numbers into a computer, and measuring how those students improve from one academic year to the next” (Hill, 2000, para. 2). While people may think it was unfair to measure scores of lower income students to those of their higher income counterparts Sanders, who agreed, said the good thing about his model is it only measured the “value added to a child’s learning” since any and all students can learn (Hill, 2000, para. 4). By analyzing the amount a student’s test score increased or

decreased, conclusions could be drawn about the performance of schools and teachers. Sanders' value-added model was touted as an "objective method" because it only examined numbers (Hill, 2000, para. 6). This comment suggested that teaching was an additive measure regardless of outside variables.

The purpose behind the value-added method was to be used as a "diagnostic" tool to help teachers improve in areas where they were lacking, which was leading to the student achievement outcome (Hill, 2000, para. 7). The idea came to Sanders in 1982 while reading an article about how teacher effectiveness could not be measured through test scores (Hill, 2000). He decided to try and prove that it could indeed be done and used a "mixed model" to evaluate farm animals (Hill, 2000, para. 11). He was given data and found a model that worked using "three years worth" of data (Hill, 2000, para. 11). The model worked by merging new test data with historical test data and tracking the student achievement; student data was compared individually and not against other students.

VAM is a covariate model. That means the model used different variables to measure student achievement. The variables accounted for in the Florida model were: historical data (up to 2 years of prior test scores), how many "subject-relevant courses" the students were enrolled in, disability status, English language learner status, gifted status, attendance, mobility, and difference from "modal" age in grade to account for grade retention (FIdoe "Florida's Value-Added Model," n.d., p. 14). Classroom characteristics used in the model are: class size and homogeneity of the students' prior test scores (FIdoe "Florida's Value-Added Model," n.d., p. 14). To continue to "level the

playing field,” the committee decided to make a teacher’s effectiveness contingent on 50% of the school characteristics as well. The rationale behind this was to keep the teacher from being held accountable for factors out of his or her control, while recognizing that the teacher contributes to the school component (Fldoe “Florida’s Value-Added Model,” n.d., p. 16).

A teacher’s VAM score is a combination of scores. The formula included the “school component” which examined the average gains made by the school in relation to the state, and a “teacher component” which examined the average gains made by the teacher’s students in relation to the school (Fldoe “Recommendations of the Florida Student Growth Implementation Committee,” n.d., p. 2). Based on this rationale, it can be speculated that a teacher who taught honors students would not have an overwhelming advantage over a teacher who taught lower level students. The formula can be summarized as: “Teacher Value Added Score=Unique Teacher Component + .50 x Common School Component, where 50% of the school component is included in the teacher’s score” (Fldoe “Recommendations of the Florida Student Growth Implementation Committee,” n.d., p. 2).

The value-added model did not focus on the test scores; it focused on how much student progress was made. For example, in a K-4 school of 500 students, school test scores were consistently high, but the value-added scores were low: Even though students were scoring high on tests, they were not making the expected yearly progress (Hill, 2000). This analysis of VAM scores has led school personnel to realize that teachers were, “...devoting more time and energy to the school’s neediest children...but not

spending as much time working with the top students,” this had caused, “...academic gains to drop off as achievement rises...” (Hill, 2000, para. 37). Even though students were achieving, gains were not evident because of the way growth was measured.

VAM controls for different student, classroom, and school characteristics in the statistical model, and by examining these different characteristics, VAM measured the “impact of a teacher on student learning” (Fl DOE “Florida’s Value-Added Model,” n.d., p. 6). The model did not evaluate teachers based on one year of scores or “growth from one year to the next” (Fl DOE “Florida’s Value-Added Model,” n.d., p. 6). The model examined how a student performed previously on a test, used the variables in an algorithm to predict the next year’s growth and then measured the actual difference in growth. If the actual performance exceeded the predicted performance then the teacher received a positive VAM score. If the actual performance was below the predicted performance then the teacher receives a negative score. If the predicted score and the actual score were the same, then the progress was simply a met expectation (Fl DOE “Florida’s Value-Added Model,” n.d.). The figure of the Value-Added Example shows how the differences in a student’s performance could create a positive, negative, or neutral score.

Value-Added Example

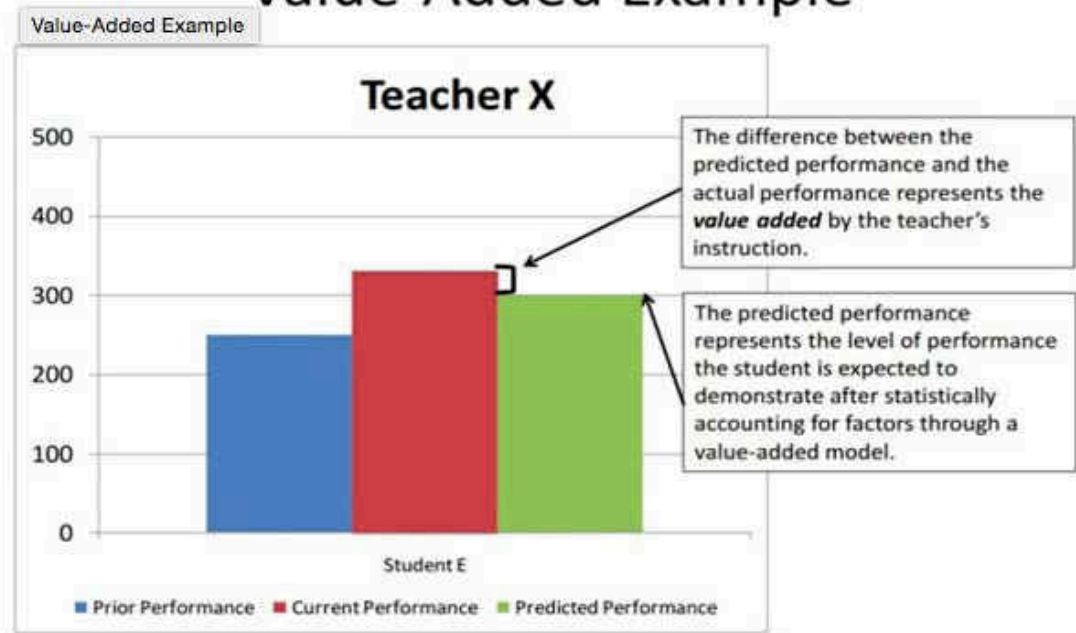


Figure 1 Difference of performance in Florida's Value-Added Model

Source: Florida Department of Education <http://www.fldoe.org/teaching/performance-evaluation>.

VAM is a statistical model that measures student growth using an algorithm. The variables used in Florida's VAM account for teacher contribution and some student contributions in the classroom. Using student assessment scores, value-added scores are calculated into a teacher's final evaluation for a school year. The VAM score used depends on the students assigned to a teacher and the corresponding assessments for the class.

Marzano

In order to receive Race to the Top money, Florida had to adopt and implement an evaluation system. Many school districts in Florida use the Marzano system. The

Marzano system was designed to improve teaching, not to be used as an accountability system, and it focuses on specific teaching strategies (Quinn, 2014). Marzano says there is a difference between observations used for evaluations and observations used for improvement, and in order to get an accurate picture of what a teacher is really doing many observations (more than four) are needed (Quinn, 2014). A difficulty with using growth models as accountability models is inflated scores. According to Marzano in Quinn's interview (2014) "...evaluators tend to inflate teachers' scores dramatically...probably because they want to give everybody the benefit of the doubt" (p. 15). When a teacher's evaluation can determine pay that may put a lot of pressure on an evaluator. By rating teachers highly the model is not working to help the teacher grow because there is no room being left for development (Quinn, 2014). Based on this information, observers tend to err on the side of the teacher, which does not help improve instructional practices but tell teachers everything they are already doing is correct.

According to Quinn's interview with Dr. Robert Marzano (2014) the Marzano model is actually an internationally used model. Based on observations, Marzano states that teachers are very similar internationally as they have some of the same concerns; "It's the systems that are so different" (Quinn, 2014, p. 16). Systems may be different internationally, but they are also different locally as different districts within the same state use diverse models for delivering feedback. Assessing administrators and observers know what to look for, but there is no systematic way that data is collected in order to control for those factors (Quinn, 2014). "The central role of leadership then, becomes to monitor what makes the bigger difference and to do so frequently" and then "taking

action right away” as effective leaders would do (p. 16). Monitoring is one of the central pieces in the Marzano evaluation system as evaluators look to see if teachers are monitor their students’ comprehension and progress. From that system, observers have the opportunity to provide feedback so teachers can improve their craft and therefore improve student achievement. Taking action may be more than simply providing feedback. Depending on the needs of the teacher, action may include specific professional development and coaching.

The Student Success Act and Florida statutes created a structure from which leaders could choose evaluation models. The Marzano model, which includes Deliberate Practice, are concepts taken from Dr. Robert Marzano’s research on instructional strategies, and satisfied statutory requirements. In Marzano’s book, *Becoming a Reflective Teacher* (2012), he discussed how teachers must practice to develop their pedagogical skills, and he stated that reflection improves teachers’ skills, which in turn improves student achievement. “A corollary is that teacher reflection improves teacher pedagogical skill” (Marzano, 2012, p. 3). Theory and research claim teacher reflection is critical to student learning, but not widely “embraced” in K-12 education (Marzano, 2012, p. 4). Since practice is so important to teachers’ skills, they must develop these skills using a “metacognitive” model called Deliberate Practice (Marzano, 2012, p.8). While it is probable that teachers do reflect on their teaching, it is not embraced in the systematic way as presented as Marzano.

According to Dr. Marzano, “Deliberate practice, identified by Ericsson and his colleagues (1993), flies in the face of what most people think of as practice” (Marzano,

2012, p. 6). It is a type of practice designed to “resist automaticity” and increase cognitive processing (Marzano, 2012, p. 7). The process helps teachers move away from routine behaviors in teaching and reflect on pedagogical practices intrinsically. Deliberate Practice also differs in it is “not inherently enjoyable” and it “also requires large amounts of time” (Marzano, 2012, p. 7). Deliberate practice requires teachers to choose one instructional strategy as a focus for development over the school year. There are 41 elements to choose from and teachers use this process as a continual self-reflection of their work in the specific area (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012). Many teachers, especially veteran teachers, may feel as though their practices do not need development because they have been teaching for many years. As quoted by Marzano, Ericsson claims this type of practice is meant to build performance “slowly over a very long time” to achieve the “highest levels of performance” (Marzano, 2012, p. 7). During the 2013-2014 school year, Deliberate Practice was mandatory for all teachers in Large Urban School District. The Deliberate Practice score was used in conjunction with the instructional practice score to make up half of a teacher’s final evaluation (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012). A score reflective of student achievement made up the other 50% of the teacher’s final evaluation score (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012). The instructional practice score was reflective of the teachers’ progress using pedagogical strategies.

To become an expert, “a teacher must have a general idea of what constitutes effective teaching” (Marzano, 2012, p. 11). Teachers need to have some aspect of

background knowledge to become experts in their respective educational fields. The “road to expertise” begins with setting goals (Marzano, 2012, p. 13). Goal setting was the self-rating portion of Deliberate Practice. The next step included picking a target element to work on throughout the year or “focused practice of specific strategies” (Marzano, 2012, p. 14). This was the element the teachers must purposefully plan on increasing correct practice through research and feedback. The observation step was called, “receiving focused feedback,” and “observing and discussing teaching” (Marzano, 2012, pp. 15-16). At this step the teacher received feedback on his or her instructional practices. The overall goal for this model was to get teachers to continuously reflect on their teaching in order to improve student achievement. The following is a look at how Marzano’s model was used in Large Urban School District and statutory requirements that each element met.

In the large urban school district of this study, procedures for personnel evaluations were “designed to contribute toward achievements of goals identified in the district’s plan pursuant to state statute” (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 4). State statute 1012.34 (1)(a) stated, “For the purpose of increasing student learning growth by improving the quality of instructional, administrative, and supervisory services in the public schools of the state, the district school superintendent shall establish procedures for evaluating the performance of duties and responsibilities of all instructional, administrative, and supervisory personnel employed by the school district” (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p.4).

Large Urban District schools used an evaluation model based on Dr. Robert Marzano's research. The district "modified the state adopted model" (Ocps "Instructional Personnel Education System Procedures Manual: 2012-2013," 2012, p. 4) that combined student growth with teacher growth. For the 2012-2013 school year, 40% of the teachers' evaluation came from VAM scores and 60% came from Marzano's model, while the 2013-2014 model was a 50/50 split between VAM and Marzano (Ocps "Instructional Personnel Education System Procedures Manual: 2012-2013," 2012). The rationale behind the Marzano model was teachers would be able to "increase their expertise" each year, and "a common language of instruction" is important to "provide a foundation of professional conversation (Ocps "Instructional Personnel Education System Procedures Manual: 2012-2013," 2012, glossary p. 2). The Marzano model provided a framework for feedback used for "teacher growth," and the model would improve "teacher efficacy," which would cause the improvement of "student learning" (Ocps "Instructional Personnel Education System Procedures Manual: 2012-2013," 2012, p. 15). The Marzano instructional framework provided teachers with opportunities to develop their pedagogical practices throughout the year each year. It also provided an avenue for conversation using common language as it related to best instructional practices. Ultimately the focus cycles back to student achievement, practice, reflective teaching, feedback, and building self-efficacy to increase student achievement.

There were four "domains" in the Marzano model and each domain was worth a certain percentage of the evaluation (Ocps "Instructional Personnel Education System Procedures Manual: 2012-2013," 2012, p. 12). "Domain 1: Classroom Strategies and

Behaviors” included information from formal and informal observations and examined what the teacher was doing in the classroom through rules and procedures, content, and student engagement, and was worth 60% of the evaluation (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p.9). “Domain 2: Planning and Preparing” was specific to lesson planning, and was worth 20% of the evaluation (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p.9). “Domain 3: Reflecting on Teaching” included teacher self-assessment and reflection of instructional strategies (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p.9). This piece focused on how the teacher tracked his or her progress using pedagogical practices.

Before each formally scheduled observation, the teacher met with their assessing administrator for a preconference and after the observation for a post-conference. There are no scheduled conferences for informal observations. A self-assessment should have been filled out before each conference. There is a form for the pre-conference, which asks what the teacher will do and a separate form for the post-conference, which inquires about what the teacher did and how the lesson went. This was worth 10% of the evaluation. The last domain, “Domain 4: Collegiality and Professionalism” was also worth 10% of the evaluation (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p.9) and focused on the conferences between the teacher and administrator and any other documented circumstances a teacher participated in or worked professionally with administrators, students, community members, or other teachers such as participation in professional learning communities.

To be an observer or evaluator, a person must have gone through “intensive training” and become certified (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 7). State statute 1012.34 (2) (f) stated, administrators must, “Provide for training programs that are based upon guidelines provided by the department to ensure that all individuals with evaluation responsibilities understand the proper use of the evaluation criteria and procedures” (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 7). In order to ensure consistent implementation of observation measures, “the district will monitor teacher evaluations for consistency between performance scores and student growth scores, and where discrepancies exist, additional training will be provided to the evaluator” (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 7). There may have been monitoring to ensure scores were not inflated for example, if a teacher was rated as highly effective but did not show student growth, or vice versa, that may have been considered a discrepancy for the district to address.

The evaluation process for each teacher was dependent upon which category a teacher was placed in based on “experience and expertise” (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 8). There were 5 categories for teacher placement: “Category 1” teachers were new to teaching, “Category 2A” teachers have had at least three years of experience, and “Category 2B” teachers have experience but may have been new to the district, “Category 3” teachers were on improvement plans for unsatisfactory evaluations, and “Category 4” teachers did not have

enough data to be placed in a different category (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, pp.8-9).

The category a teacher was placed in was significant because it determined how many observations a teacher received. As stated, these categories varied depending on multiple factors such as experience, subjects taught, and previous performance. Category 1 and 2B teachers received 2 formal observations and 4 informal observations. Category 2A teachers received 1 formal and 2 informal observations. Category 3 teachers received 3 formal and 7 informal observations (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 11). These numbers were the “minimum number of formal and informal observations required for each category” (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 11).

There were two types of evaluations: formal and informal. The informal observation could have been “announced” or “unannounced” (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 11). The idea was not to try and catch the teacher doing something wrong, but to get an accurate picture of the teacher’s classroom and then be able to give the teacher feedback. The first informal may or may not be used as part of the evaluation score. For the informal observation, an observer should have been in the classroom for at least 10 minutes, and did not require any planning conferences. Informal observations provided a platform to offer feedback to teachers, which should be delivered in a timely manner (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012).

Formal observations were scheduled observations. Florida statute 1012.34 (3) (b) stated, “All personnel must be fully informed of the criteria and procedures associated with the evaluation process before the evaluation takes place.” This process required meetings with the assessing administrator that should have lasted a minimum of 30 minutes. This process did not determine the teacher’s final rating, but the formal observation was used as a primary source of data for the final effectiveness rating. The formal observation allowed teachers to “reflect upon” their pedagogical practices (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 11).

Teachers were scored on a “5-level rubric” from 0-4 (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012, p. 12). The rubric was interpreted as follows: 0 meant the teacher was not using a skill, 1 meant the teacher had a beginning concept of the skill and needed further work, 2 meant the teacher was developing his or her skill and performed most of it correctly with parts missing, 3 meant the teacher was successfully applying a skill, and 4 meant the teacher was innovating at teaching a skill by adjusting instruction to meet the needs of all students (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012). The skill an evaluator was looking for was based upon the teacher’s learning goals for that lesson. A teacher may request additional evaluations, and if a teacher had more evaluations than needed, their lowest score was dropped from their final evaluation score (Ocps “Instructional Personnel Education System Procedures Manual: 2012-2013,” 2012).

The Marzano model is a tool designed to develop teachers' pedagogical practices. Many counties, including Large Urban School District in Florida used it as part of the evaluation model adopted as a result of state statute and Race to the Top. Teaching experience within the school district dictated how many formal and informal observations a teacher received. While the processes may not be enjoyable for teachers, the goal of observations was to develop practice, raise efficacy, and increase student achievement.

Supervision

In schools there was confusion over supervision and evaluation in terms of what it meant to be a supervisor and how that related to helping teachers improve. According to Tang & Chow (2006) "Supervision in the form of lesson observation and post-observation conference and the communication of constructive feedback in supervisory conferences is essential to teachers' professional development" (p. 1066). Supervision is not only about evaluating teachers but also about developing them through observations and providing feedback.

Evaluation may be thought of as punitive by having an effect on pay and job position. In schools, supervisors were tasked with performing observations and evaluations. The defining role of supervisor falls to the discretion of the school leadership. Perceptions on the role of the supervisor and how it related to evaluation originated from the changing role of the supervisor throughout history. In the late 1800s, school quality meant supervisors would get rid of what were thought as "incompetent

teachers” (Nolan & Hoover, 2011, p. 3). Research by Frase (2001) has revealed that within schools, teachers admitted some of their colleagues were judged as “incompetent,” but their evaluations showed quite the opposite (p. 176). This implied observation scores were inflated. Potential reasons for inflation of scores could include evaluators who erred on the side of the teacher because of the implications of an unsatisfactory evaluation, maintaining the status quo of a school environment, or misunderstanding of the evaluation scoring process.

Different philosophies on supervision included: social efficiency, which related to bureaucracy and teacher rating scales; progressive supervision, which emphasized collaboration and the importance of the teachers; and democratic supervision, which emphasized the flexibility of the organization (Nolan and Hoover, 2011). Different time periods saw the predominance of certain models. After the launching of Sputnik, an increase in the role of the supervisor as evaluator saw the use of teacher evaluations as punitive measures if a teacher fell into noncompliance (Nolan & Hoover, 2011). Kelting, Jenkins, & Gaudreault (2014) discussed the purpose of clinical supervision as developing teachers through feedback to improve instruction, and not used as a punitive or evaluative tool.

Nolan and Hoover 2011 discussed supervision and evaluation as two separate functions that could work together. Teachers in Large Urban School District in Florida viewed these systems as analogous because formative supervision counted toward the final evaluation. Frase (2001) says teachers have the right to dislike evaluation and supervision because historically they have not been helpful due to lack of constructive

feedback. Formative supervision, or the informal evaluation or walkthrough as used in Large Urban School District, did utilize the clinical model of pre and post observation conferences for informal evaluations. While the clinical supervision model consisted of conferences before and after observations for development purposes (Range, Young, & Hvidston, 2013; Kelting, Jenkins, & Gaudreault, 2014; Tang & Chow, 2006), this process was only required for formal evaluations in Large Urban School District. Informal evaluations occurred within a specific timeframe and were not scheduled observations. An administrator may walk in and observe at any time and usually deliver feedback electronically through iObservation, which was the online location for teacher growth plans and observations.

According to Range, Young, & Hvidston (2013) formative supervision is an opportunity to coach teachers and build capacity in their instruction (p. 62). In Large Urban School District in Florida personal, face-to-face feedback only really occurred during the formal evaluation process. Clinical supervision requires time, which can be a difficult commodity to find as an administrator. Often administrators are taken away from instructional supervision to work with managerial duties and issues. Some schools in Large Urban School District, Florida required their instructional deans and school based resource teachers or coaches to observe and provide feedback to teachers. Unfortunately, these observations lacked the pre and post observation conferencing and delivered the message that those who were supposed to coach or support teachers were now evaluative figures.

Supervisors may be thought of as administrators because the connotation of supervisor suggests someone who has authority over another person. Nolan & Hoover (2011) discussed how supervisory roles could include peers such as coaches or even department chairs. In schools, curriculum resource teachers and content area coaches may be instructional personnel who hold no authority over their peer teachers, but hold supervisory duties such as helping to develop their peers through professional learning, training, and coaching. Nolan & Hoover (2011) noted that “Although most supervisors would like to work with teachers in helpful and supportive ways, they have also been saddled with the duties of teacher evaluation...the vast majority of supervisors have become de facto teacher evaluators who do little or no supervision” (p. 19).

The distinction between evaluation and supervision could be clearly made between who was simply supervising teacher learning and growth and who is evaluating the teacher’s performance. In clinical supervision models, the role of supervisor and evaluator should not be the same as teachers may view supervisory practices as evaluative in nature (Nolan & Hoover, 2011). While the scope of evaluation consisted of a broad look at how a teacher performed throughout the year, supervision was supposed to have a narrow focus on specific elements in which teachers may have needed help or support (Nolan & Hoover, 2011). Supervisory practices were ideal for feedback and development of teachers.

The nature of the relationship between the teacher and those supervising or evaluating was very different. The role of the evaluator was generally not one of closeness and trust because of the implications of evaluation. Supervisors, such as

coaches, should have been able to establish trusting relationships with teachers and create a culture where the teacher sought feedback and development knowing that changing behaviors and having honest conversations would not negatively impact the teacher's employment or pay. "The supervisor should be on the teacher's side in helping the teacher to accomplish mutually set goals and objectives" (Nolan & Hoover, 2011, p. 10). A noted difference between evaluations and supervision was evaluative observations were announced and unannounced visits while clinical supervision began with a preconference to discuss teaching and the visit did not count toward a summative evaluation (Nolan & Hoover, 2011).

An important aspect to evaluation was a teacher needed to see the value and worth of the evaluation system, "When teachers do not buy into the system, they can find subtle ways of undermining the goals that the system is attempting to achieve" (Nolan & Hoover, 2011, p. 15). If teachers did not accept and own the evaluation system, then the only reason for them to attempt to apply any of the research-based decisions districts made to improve student learning was compliance. An example of this compliance was related to the Marzano framework in which an element required teachers to post learning goals and scales for students so they knew what they are learning. Research showed that there was a .56 effect size for learning goals and scales as they related to student achievement (Hattie, 2009). Teachers needed training and deep understanding of the evaluation models so they were not posting meaningless goals and scales just in case an administrator walked in, but posting information so students were aware of what they would be learning and could monitor their own progress as they learned.

Evaluation was a summative measurement of a teacher's competence and performance while supervision was aimed at helping teachers grow through intrinsic motivation to improve (Nolan & Hoover, 2011). Teachers who were able to take ownership of their own growth leads back to Bandura's (1977) theory of self-efficacy: To be able to carry out a task he or she believed they were capable of completing. In some cases, since growth cannot be forced supervision may not have been enough to help teachers without the added motivation of evaluation output, while other teachers strived to improve because they truly wanted to do what was best for students.

There were distinct differences between supervision and evaluation. To develop teachers, they needed supervision from a trusted supervisor. Many administrators may have found it difficult to be effective supervisors and evaluators simultaneously. Supervisors were often used to evaluate, which caused further frustrations toward developmental methods and models used in schools. To be receptive to feedback, supervision is needed to lead to building capacity not evaluation.

Feedback

“Communicating feedback productively is of paramount importance to professional learning in many professions” (Tang & Chow, 2006, p. 1066). In addition as an accountability tool, teacher evaluation was used for the purpose of monitoring teacher practice in order for administrators to give feedback for improvement. According to Hill & Grossman (2013) policy makers believed feedback and professional development for teachers were the largest intents of evaluation systems. Although evaluation systems

may have been designed to improve instruction, Hill & Grossman (2013) argued the current design would cause the system to fail. In order for the current system to be effective, teacher evaluation designs should be subject-specific, scored by content experts, and include adequate and usable feedback (Hill & Grossman, 2013). These three items work in tandem to improve teacher learning and are related to giving teachers the development needed to improve instruction. Marzano (2012) stated, "...a teacher's pedagogical skill in the classroom is causally linked with how well and how much students learn" (p. 3). In order for teachers to improve their practice and become more reflective, feedback is required to help teachers focus on specific skills. Feedback has an effect size of 0.73, which means there is a probability that 73% of students will increase in achievement compared to those who do not receive feedback (Hattie, 2009).

With current evaluation systems generalized to cover all grade levels and subjects, the absence of subject specific tools leaves the teacher without subject-specific feedback. While a generic tool could be used to monitor teaching strategies, it did not take into effect specific, complex, and higher-order distinctions in each subject area or grade level (Hill & Grossman, 2013). In order to help teachers improve their instructional practices, subject-specific information should be monitored by content experts. Often those observing teachers were not knowledgeable in content across the curriculum. Hill and Grossman (2013) argued that raters should be content experts in order to give more effective feedback. Content-specific knowledge of the raters would allow feedback to be focused on the use of best practices as they related to the subject as opposed to generic instructional practices.

While evaluation was also a tool for helping teachers improve their pedagogical practices, the focus on accountability may have made teachers feel like it was purely a tool used to monitor quality and competency. Evaluation was an avenue where school leaders could deliver feedback that was individualized to what the teacher and students needed. Research suggested, "...there is an absence of systematic feedback for teachers to facilitate their professional growth and improve instruction" (Anast-May, Penick, Schroyer, & Howell, 2011, p. 2). Feedback given to teachers during observation and evaluation should not only be about what the supervisor saw, but also specifically what a teacher could do to improve. This feedback should also be given in a timely manner so it is fresh and relevant to the corresponding lesson. Kelting, Jenkins, & Gaudreault (2014) recommend teachers receive feedback using videotaped lessons within 24 hours or less of the observation. Most teachers who did not have the added benefit of video recall would probably also benefit from immediate feedback. Giving feedback should not only be a documented retelling of the lesson observed, it should include giving suggestions on types of development and learning a teacher could participate in order to meet and exceed the quality assurance expectations in an evaluation.

In order to have reliability in observation scores, one rater must observe at least three to four times or multiple raters should observe a specific teacher (Hill & Grossman, 2013). Accuracy was important to feedback and improving teaching quality. If a rater missed content inaccuracies during a lesson then that was a missed opportunity for feedback and improvement (Hill & Grossman, 2013). Since many raters were principals or assistant principals, time was an issue for feedback and follow-up for teachers. In

order for feedback to be more successful, research on coaching has shown that it must be “specific” and “actionable” in relation to the teacher’s content area skills (Hill & Grossman, 2013, p. 379). If feedback was given, it was important for the rater to follow-up with the teacher to discuss what feedback was implemented and the outcomes. Since time was an issue, Hill & Grossman (2013), discussed the concept of observing teachers and giving feedback to those who have a genuine need or were struggling in the classroom instead of trying to observe everyone for the sake of accountability. To improve student learning, it is important to foster teacher learning through purposeful, actionable feedback, follow-up, and professional development. Using evaluation for the sole purpose of holding teachers accountable for test scores will not produce student results (Hill & Grossman, 2013).

Feedback is a powerful tool on learning, but the timing, type, and delivery have different effects. Feedback can be defined as “...information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding; a “consequence” of performance” (Hattie & Timperley, 2007, p. 81). In order to be effective, feedback must have been related to something known and given within a context (Hattie & Timperley, 2007). Research has shown the types of feedback given have made a difference. Higher effect sizes were related to “...students receiving information feedback about a task and how to do it more effectively” rather than feedback related to “...praise rewards, and punishment” (Hattie & Timperley, 2007, p. 84). While Hattie & Timperley (2007) were referring to students in this statement, any

learner who was trying to become more effective at a task could benefit from specific feedback.

Formal feedback has generally been given electronically through the iObservation system used in Florida Large Urban School District. The feedback was recorded by the observer and then sent to the teacher being observed. This process occurred whenever the rater recorded an observation for the teacher. The option of giving comments was always available regardless of the type of observation. The feedback stayed in this system and remained available for teachers to reflect on, ask questions about, and refer back to as needed. Face-to-face feedback was generally given in addition to electronic feedback when the observers and teachers met for the post-observation conference. What the rater recorded as feedback was what this study sought to determine as effective or ineffective on raising student achievement outcomes. Formative supervision was “The primary method by which principals impact teachers’ instructional practice” (Range, Young, & Hvidston, 2013, p. 61). Formative supervision was considered “proactive” because it gave teachers feedback on their instructional practice and addressed weaknesses to be fixed before teachers received summative evaluations and functioned under the “...assumption that continuous improvement is necessary for teacher growth” (Range, Young, & Hvidston, 2013, p. 61). Research has shown that constructive feedback was rarely given if and when principals were able to observe classrooms (Frase, 2001).

Feedback was more effective when it built on previous changes, was related to specific goals, and when it was delivered in low threat environments (Hattie & Timperley, 2007). Low threat environments allowed recipients to pay closer attention to

feedback and self-esteem was not threatened (Hattie & Timperley, 2007). For example, a low threat environment where teachers feel most comfortable may be the teacher's classroom. When giving feedback to teachers, it may have made a difference if feedback was delivered by a coach or administrator, informally versus formally, and evaluative versus non-evaluative. Feedback from administration carried the risk of being related to evaluation, and feedback that was given without being related to evaluation allowed for teachers to make adjustments to teaching strategies before formal observations.

According to Hattie (2009) the effect size of the teacher as related to student achievement was 0.44. This means that 44% of students will increase in achievement compared to students who do not have a quality teacher (Hattie, 2009). As stated by Range, Duncan, & Hvidston (2013), "Researchers have linked effective teaching to increased student achievement" (p. 43). Hattie (2009) stated, a quality teacher "...as rated by students, are those who challenge, who have high expectations, who encourage the study of their subject, and who value surface and deep aspects of their subject" (Hattie, 2009, p. 116). Effective teaching could be supported and developed through the amalgamation of effective supervision, coaching, feedback, and professional development.

When discussing how feedback affected student performance when the teachers were providing it, Hattie (2009) said, "Feedback is more effective when there are perceived low rather than high levels of threat to self-esteem, presumably because low threat conditions allow attention to be paid to the feedback" (p. 175). Based on Bandura's (1977) theory of self-efficacy, if the person who received the feedback had

high self-efficacy then the suggestions could be more easily accepted and put into practice. If feedback was given to someone with low self-efficacy, then the suggestions may have become more about the person and not about the practice, which could make the feedback ineffective.

Hattie (2009) stated the purpose of feedback was “To reduce discrepancies between current understandings and performance and a learning intention or goal” (p. 175). To be effective, feedback should answer three questions and work at four levels (Hattie, 2009). These three questions were: “Where am I going? (learning intentions/goals/success criteria), How am I going (self-assessment and self-evaluation), and Where to next (progression, new goals)” (Hattie, 2009, p. 177). The four levels where feedback should work were: “Task level”, “Process level”, “Self-regulation level”, and “Self level” (Hattie, 2009, p. 177). Each level of feedback had an intention to help the learner by directing toward what needs to be accomplished. At the task level feedback could include directions or more information; the process level was directed at the how the task could be understood and completed; and self-regulation feedback could require learners to reflect on what they already knew and how it could be used to enhance the task (Hattie, 2009).

Based on this model, an example of how feedback could be delivered could contain the following elements: I see you already know/can do _____. In order to achieve the goal of _____, you must/try _____. This sentence stem included the first three levels of feedback. It began by addressing the self-regulation level, then the task level and the process level. While giving praise at the self-level may

be incorporated into feedback, it was not a recommended practice to give praise alone (Hattie, 2009). According to Hattie (2009) praise alone was not effective since it did not address the three feedback questions and was generally not related to the performance on the task. While feedback should not include praise alone, it should "...provide affirmation for positive teaching characteristics..." (Range, Duncan, & Hvidston, 2013, p. 45) as teachers may feel motivated by knowing they were doing well which could increase self-efficacy.

Feedback, which included praise and the elements outlined by Hattie, (2009) could be phrased in the following way based on the focus statement and desired effect from element 1 of the Marzano protocol: I see you have a progressive, student-friendly learning goal and scale posted. Think about how you could relate the lesson back to the learning goal and scale so students know what they mean and could effectively monitor their own progress. In order to achieve applying on this element, make sure you monitor 51% or more students for the desired effect. One strategy you could consider using is an exit slip that students must show you before they leave for the day.

In order for feedback to be effective, it must be timely, proactive, clear, and given in an environment where teachers do not feel threatened. The purpose of feedback was not to highlight everything a person did incorrectly, but to help the individual be more reflective in their practices and strive to develop his or her skills. This can be achieved through including specific elements while giving feedback to clarify the intended direction of the feedback and appeal to levels as related to the task and individual. Figure

2 illustrates the feedback model by Hattie (2009) showing the purpose of feedback at the beginning to how the feedback works at the four levels.

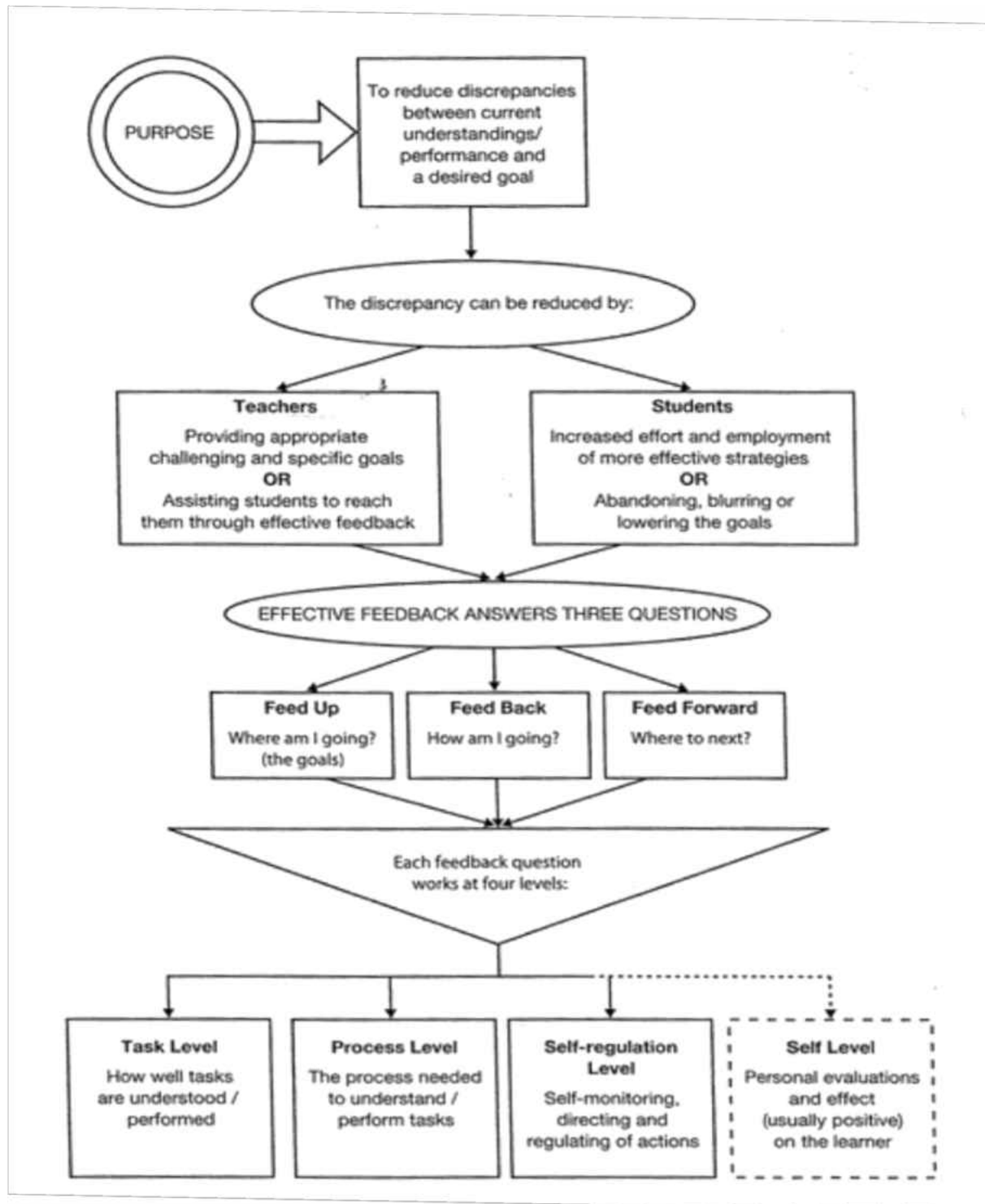


Figure 2 “A model of feedback”

Source: Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London, UK: Routledge. (p. 176).

Coaching

Evaluation should not only hold teachers accountable, but it should also seek to help teachers develop their classroom teaching and practices (Tutyens & Devos, 2013, p. 511). In order to help develop teachers, quality feedback was necessary to help direct teachers on what was specifically needed to improve instructional practices. Research has shown that teachers needed a specific type of quality feedback that was individualized in order to be successful; however, they were not receiving what they needed (Anast-May, Penick, Schroyer, & Howell, 2011). Teachers needed “...constructive and meaningful feedback” which would “...ultimately lead to an increased sense of self-efficacy in their teaching” (Anast-May, Penick, Schroyer, & Howell, 2011, p. 2). Relating feedback to Bandura’s (1977) theory of self-efficacy, teachers needed to be directed through feedback in a way that bolstered their sense of self and feelings that they were capable of putting into practice what their supervising administrators were expecting to see in the classrooms as well as positively impact “...teachers’ knowledge and skill” (Marzano & Simms, 2013).

Coaching was a method used to transform teachers through feedback, development, collaboration, and support. Marzano & Simms, (2013) while exploring multiple definitions of coaching, cited Gauthier & Giber (2006) who defined coaching as “...helping someone move from where he or she is to where he or she needs or wants to be” (p. 4). Coaches help teachers by supporting the transfer of knowledge from professional development to practice through observation, modeling, and feedback (Marzano & Simms, 2013). While coaches may be considered supervisors, they were

generally not evaluative although, some schools may have trained their coaches as evaluators and used them for classroom walkthroughs or informal evaluation purposes. In the clinical model of evaluation, it is important to have a preconference with the teacher, conduct an observation of the teacher, and provide feedback to the teacher through a post observation conference (Marzano & Simms, 2013; Nolan & Hoover, 2011; Range, Young, & Hvidston, 2013).

While coaching, giving feedback should not consist of telling the teacher what was observed or done incorrectly, or telling the teacher what he or she needs to do. Coaches should support teachers to become more reflective of their practices by evaluating their own performance in the classroom, discovering ways to improve their pedagogical practices, and determine the success of new practices (Marzano & Simms, 2013). Marzano & Simms (2013) cited research from Joyce and Showers (2002) who said coached teachers generally understood new strategies better and were able to apply and transfer that knowledge to their classrooms. There was mixed research on the effects of coaching on student achievement. Some research noted no significant effects while other research showed "...significant increases in student achievement" (Marzano & Simms, 2013, p. 7). Marzano (2013) suggested although evidence was mixed, there was more research supporting the positive effects coaching has on student achievement.

To help teachers become more reflective practitioners, coaches may provide teachers with feedback, but a number of factors may determine if the feedback was accepted and applied in classrooms. A teacher must feel as if he or she is in a nonthreatening environment, meaning the feedback was for the purpose of improvement

and not evaluation (Hattie & Timperley, 2007; Marzano & Simms, 2013). The teacher must also be willing to accept the feedback. According to Marzano & Simms (2013) “It stands to reason that if a teacher doesn’t know what he or she is doing right or wrong, it will be difficult for that teacher to improve his or her knowledge and skill” (p. 10).

Coaches as instructional personnel could help their colleagues identify areas of strength and needs for support through self-reflection and feedback.

In order to provide teachers with content-specific feedback, coaching was used to help build capacity in teachers. Some schools have subject-specific coaches such as literacy and mathematics coaches who were used to observe teachers in those content areas (Hill & Grossman, 2013). In order for coaching to be successful, feedback must be non-evaluative in nature, private, and mentees must feel like they were in a non-threatening environment (Marzano & Simms, 2013). If administrators planned to act in a coaching capacity, then it must be made clear that the purpose was for teaching and learning and not for accountability (Marzano & Simms, 2013, p. 9).

More important than who was coaching, is the ability for the person being coached to be open and accepting to feedback (Marzano & Simms, 2013, p. 9). If a person did not want to be coached or was not comfortable being coached, then the purpose of providing the feedback may not be effective. As stated by Marzano & Simms (2013) “Effective feedback should specify which strategies a teacher is performing correctly and effectively, which strategies a teacher is using but with errors or omissions, and which strategies a teacher could or should use but isn’t,” and be “timely and specific” (pp. 11 & 217). There are both formal and informal types of feedback. Informal written

feedback is brief, it may be in the form of an email or note and usually highlighted something positive observed by the coach (Marzano & Simms, 2013). Formal written feedback is more “detailed and specific” and gave the information about what the teacher did that worked well or needed to adjust (Marzano & Simms, 2013, p. 218).

In examining research on coaching, Reeves (2009) cited differing opinions on what coaching is and should be. Reeves (2009) cited Luecke (2004) who said “...managers and supervisors should use coaching strategies with their direct reports in order to improve performance and teamwork” (p. 72). An example of this in education would consist of observers giving coaching tips and feedback to teachers when visiting classrooms. Reeves (2009) also cited Goldsmith & Lyons (2006) with differing opinions that suggested “...the coaching relationship is distinct from supervision and evaluation and recommends that the role of coach be assigned to an independent person not in the direct line of supervision of the person receiving coaching” (p. 72). This suggestion limited who could coach and who could give feedback that counted toward an evaluation. An administrator should not try to be a coach if he or she evaluates and a school-based coach should not be associated with evaluation.

Wade (1984) drew conclusions on the mere moderate effects of coaching. Having a school-based coach to monitor and follow up on professional development trainings could help determine information on behavior and results (Wade, 1984). Wade (1984) stated that evidence showed coaching did not really change teacher behavior or transfer knowledge more than moderately. The reasons behind coaching only having a moderate effect on teacher behavior comes down to research on relationships between the coach

and staff. Based on additional research, Wade noted that coaches and staff needed time to build trusting relationships where there was no worry of information being used against anyone and the development could be reflective to build independence in the teacher (1984, p. 54). Range, Duncan, & Hvidston (2013) discussed the need for administrators to build trust with teachers in order to open and encourage teachers to become reflective practitioners. Coaching and feedback work in tandem. Coaches deliver feedback and are responsible for using observational data to help teachers understand how to use feedback to improve instructional practices. Coaching is also a form of professional development. Coaches can develop teachers and provide specific trainings based on the needs of teachers.

Coaching as a form of professional development has shown mixed results. According to Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos (2009) there was research that supported coaching to help increase the efficacy of the teacher and transfer teaching practices, which did not necessarily make teachers more effective. The variability in this research may have been a result of the experience, knowledge, and practices of the coaches and how coaching as professional development was implemented (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). Mentoring and induction for new teachers have shown to be effective methods of professional development especially for new teachers where mentors could observe teachers and give collegial feedback related to what was relevant to the teaching practices (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). A “...requirement for successful coaching is feedback that is specific, accurate, and timely” (Reeves, 2009, p.

75). Coaches are not just individuals to listen to problems and complaints or provide professional development; coaches are professionals who have a wide lens to focus not only on the teacher, but the needs of the school and district as well (Reeves, 2009).

Professional Development

While professional development is considered a best practice to keep educators current on teaching and learning, it was also a statutory requirement in Florida. “The purpose of the professional development system was to increase student achievement, enhance classroom instructional strategies that promoted rigor and relevance throughout the curriculum, and prepared students for continuing education and the workforce” (Statute 1012.98). Accountability for professional development fell upon the school districts as they were required to create their own professional development system, which was evaluated by the Florida Department of Education (Fldoe “Professional Development Standards,” 2015). While research has shown that some school systems were building professional development that was ongoing and continuous for teachers, Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos (2009) “...found that such well-designed professional development is still relatively rare, and few of the nation’s teachers have access to regular opportunities for intensive learning (p. 19).

Teachers in Florida must also submit a certain number of inservice hours that could be obtained through college education courses or professional development in order to renew certification (Fldoe, “Florida Educator Certification Renewal Requirements,” 2015). This illustrates the importance of teacher professional

development. The NCLB act required schools receiving Title I funds to spend a specific amount on professional development aimed at addressing the academic concern (and to raise teacher quality) (Usdoe, “No Child Left Behind,” 2002). Race to the Top had criteria for professional development to be used to help educators transition to using new standards and assessments, to use data to drive instruction, to improve teacher effectiveness, and to improve support (Usdoe, “Race to the Top Executive Summary,” 2009).

Reviews on professional development have typically shown the changes in teachers and not how the development affected student output (Hattie, 2009). The effect size for professional development as recorded by Hattie (2009) was 0.62, a relatively high effect. Professional development is critical component in maintaining teacher effectiveness. Teachers need development on pedagogical, content, technology, management, and evaluation information. Research has shown that there were specific themes that made professional development more effective such as engaging teachers in their own learning (Kwakman, 2003; Hattie, 2009; Muijs, Kyriakides, van der Werf, Creemers, Timperley, & Earl, 2014), professional learning communities that functioned in tandem with other elements of teacher learning, and relevant, deliberate, and consistent learning opportunities supported by school leadership (Hattie, 2009; Muijs et al., 2014). Development should also be an ongoing process or occur over time (Nolan & Hoover, 2011; Hattie, 2009; Muijs et al., 2014). According to Nolan and Hoover (2011) as teachers come and go ongoing development would ensure that new personnel received training.

Research has shown that a problem with teacher evaluation was that, “Professional development was not in line with teachers’ needs identified through teacher evaluation” and that development activities may not have even been aligned with what a teacher’s needed to implement in his or her classroom (Tuytens & Devos, 2013, p. 509). According to Reeves (2009) those in education receive plenty of professional development, but typically what was learned was not put into practice. The answer is not giving more professional development, but identifying areas of need and providing development to teachers based on what they desired to learn and what they needed to develop.

Cameron, Mulholland, & Branson (2013) identified variables to consider for teacher professional development such as the professional needs of the teacher, what school systems required of teachers, and the learning needs of teachers. Just like student learning, teacher learning is not a one-size-fits-all model. Professional development should be targeted to improving specific pedagogical needs in order for teachers to apply what they have learned. If a teacher could not find the value or acceptance of professional development, then they may not have used the behaviors and strategies with their students at all. Darling Hammond (1998) stated, “Growing evidence suggests that ongoing professional development not only makes teachers feel better about their practice, but it also reaps learning gains for students” (p. 10). While professional development was designed to enhance teacher learning, the purpose of delivering professional development was to help teachers help their students.

Professional development could be given through classes, mentoring, and coaching, but “Teachers learn best by studying, doing, and reflecting; by collaborating with other teachers; by looking closely at students and their work; and by sharing what they see” (Darling-Hammond, 1998, p. 8). This type of development could be achieved through instructional rounds in which teachers watch their colleagues teaching and then reflect on what could be used in their own classrooms; a non-evaluative procedure. Teachers could also receive this type of development during their professional learning community meetings with their colleagues. In these meetings, discussing student data, pedagogical practices, assessment, and reflection could be key in developing teachers both individually and as a team. Development for teachers cannot just be about delivering a lesson; teachers need to be engaged in their professional learning just as students need to be engaged in their classroom learning. Teachers need to be involved in self-directed and differentiated learning (Kwakman, 2003; Hoekstra, 2009).

According to Muijs et al. (2014) development begins with setting student goals and determining what worked and what did not work for students to achieve those goals. Reflective instruction helps teachers question the skills they already have and what they need in order to fill the gaps to improve instruction and engagement for their students (Marzano 2012; Muijs et al., 2014). Often this reflection needs to be assessed not only by the teacher, but by someone who could help the teacher reflect; however, teachers must be involved in determining the development needed as external opinions may cause resistance and only superficial changes may be made where deeper changes are needed (Muijs et al., 2014).

Historically, most research on staff development has evaluated satisfaction of trainings rather than how behavior was changed (Wade, 1984). In order to keep teachers abreast of effective instructional strategies to increase student achievement, teachers need professional development. Professional development must also fit the needs of the type of teacher. A novice teacher may need more and different types of professional development than a veteran teacher. According to Cameron, Mulholland, & Branson (2013) veteran teachers were generally more selective when choosing professional development; however, "...school systems, their accrediting bodies, and school administrators need to communicate and demonstrate to teachers that learning in community is valued and rewarded" (p. 383). Regardless of the experience of a teacher, professional development is important and needs to be supported and valued as a school and through leadership. While collegiality and professionalism was a domain on the Marzano protocol for the large urban school district of this study, the percentage towards the final evaluation was not as large as the content section (10% vs. 30%) (Ocps "Instructional Personnel Education System Procedures Manual: 2012-2013," 2012). Professional development should mostly impact how content is delivered as long as it is monitored and carried out systematically.

A meta-analysis by Ruth Wade (1984) showed how professional development was measured using effect sizes and variables that contributed to higher effects. According to this research, inservice or professional development programs averaged a moderate effect size of .52 (Wade, 1984, p. 50). The results were viewed through four different categories: "reaction," "learning," "behavior," and "results" (Wade, 1984, p.

50). Reaction determined how participants felt about the training; learning determined how much participants learned; behavior measured whether a participant changed his or her practices as a result of the training; and the results determined if the training had any effect on the outcomes of students (Wade, 1984).

In research of program evaluation models such as the Kirkpatrick model, two levels in the model were difficult to assess: behavior and results (Owston, 2008). Behavior was difficult to assess because one would have to observe behavioral and knowledge changes within a participant who attended trainings (Owston, 2008). Results could be difficult to assess long term once a participant has moved past the confines of the organization in which he or she received training or once those who benefited from the participant's training knowledge have moved on from the confines of the organization, for example if a teacher leaves or a student graduates (Owston, 2008). Results could be formatively assessed in education with test scores, but once students graduate or leave school, results may be more difficult to track. This model was also originally for corporate trainings, and other models have been influenced by Kirkpatrick's design specifically for professional development in education (Owston, 2008).

Collecting surface satisfaction data may not have been difficult, but following up on behavior changes with specific teachers as a result of trainings could only really be accomplished through observation. Student results could also only really be monitored quantitatively through specific assessments and it could be difficult to draw a conclusion that results changed as a direct cause of the professional development. Wade's (1984)

research on professional development was summarized by the following six outcomes: 1) elementary and secondary teachers should train together; 2) there should be involvement in state, federal, or university programs; 3) incentives to participate should be offered; 4) independent study should be encouraged; 5) instructors should take responsibility for setting goals and teaching the class, and 6) observations, audio and video feedback, and practice should be considered among instructional techniques instead of pure lecture (p. 53).

Borko (2004) cited a report by The Teaching Commission (2004) which said teachers needed an approach "...that includes high standards for teacher classroom performance and student achievement, and 'ongoing and targeted professional development' to help teachers meet the demanding new standards" (p. 3). Professional development should occur continuously and be specific to the needs of the teacher. According to Borko (2004) professional development that was offered did not generally meet the needs of teachers even though the need had been identified and large amounts of money have been spent on professional development activities.

"Rigorous research suggested that sustained and intensive professional learning for teachers was related to student-achievement gains" (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009, p. 9). With new standards, teachers need development to make sure they are transferring the knowledge students need to be successful (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). Teachers need learning that is continuous and relevant to be successful. "In an effective professional learning system, school leaders learn from experts, mentors, and their peers

about how to become true instructional leaders” (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009, p. 3). Professional development must be a school-wide effort in which teachers work together collegially and leaders support teachers.

Professional development is not only workshops and trainings where consultants are paid to teach educators about programs and paradigm shifts; professional development includes collaboration between educators that occurs authentically and on-the-job. Hoekstra (2009) concluded “Our findings confirm that what teachers themselves report to be activities in the workplace they learn from, for a number of teachers indeed seem to contribute to a change in conceptions or behavior” (p. 672). Professional learning communities, collegial conversations, coaching, mentoring, instructional rounds/observations, school visits, university courses, and trainings are examples of professional development that could contribute to changing teacher practices in the classroom.

An example of data-driven professional development is teacher led professional learning communities (PLCs). In these PLCs, teachers discuss data and collaborate with other teachers to determine areas that need growth, share pedagogical practices to help them achieve results, and set goals for future meetings (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). Trends in United States teachers’ participation in professional development included teachers not receiving enough relevant professional development where they were able to interact deeply with content and collaboratively with peers (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009).

Teachers need to both know about professional development opportunities and be given the time to participate in professional development that they find enriching in order to change behaviors and transfer knowledge to students. School leadership should be supportive in providing teachers with professional development opportunities needed.

Research has shown secondary teachers participated more in out of school trainings than elementary teachers, and teachers in higher need schools received more professional development (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). Urban schools showed higher levels of participation in professional development, especially since NCLB provided funding for professional development, but there were mixed results in the levels of support in urban schools that fluctuated based on the minority enrollment (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). Higher minority enrollment showed more supports for traditional professional development outside of the school time and lower minority enrollment showed higher supports for professional development that occurred during contract time (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). The use of “release time or stipends” (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009) suggested professional development consisted mostly of out of school trainings. Other school supports for professional development also needed to consist of administrators and supervisors who lead teachers to relevant professional development opportunities based on knowledge of state and district supported trainings, as well as knowledge of which educators could help within the school.

Research in a study by Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos (2009) showed teachers reported participating in “regularly scheduled collaboration with other teachers on issues of instruction” (p. 23) there was no evidence within that report that the meetings were enriching and relevant to teacher needs or that they were in addition to what may be school-wide mandated meetings. “Again, these findings suggested that, “...the kind of job-embedded collaborative learning that has been found to be important in promoting instructional improvement and student achievement is not a common feature of professional development across many schools” (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009, p. 25). The figure below illustrates how professional development is cyclical. The reason for professional development is to support teacher learning for the purpose of increasing student achievement. The figure shows how professional development begins and ends with the student. Adjustments are made based on the needs of students and teachers and what impacted student achievement.



Figure 3 “*Teacher enquiry and knowledge-building cycle to promote important outcomes for students.*”

Source: Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State Of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*. (p. 247).

Summary

This study was conducted to determine the relationship between the content of specific-targeted feedback given to teachers through observations and student achievement outcomes as measured by FCAT 2.0. This purpose of this review of literature was to provide information on the research of this study. This review included background information, historical timelines, and reviews of empirical studies to clarify the topics in this study.

Throughout this literature review topics consistently referenced other topics. While accountability, assessment, evaluation, and supervision were all independent variables; they were not mutually exclusive and in fact, functioned together.

Literature was purposefully reviewed to support the research questions of this study, relate back to the theoretical framework of self-efficacy, and addressed 3 specific topics with related subtopics: (a) accountability and assessment, (b) evaluation, and (c) supervision. The next chapter, methodology, describes how the research was conducted from obtaining permissions to procedures on gathering and analyzing data. The methods, data sorting, coding, and tests used will be further described in chapter 3.

CHAPTER 3 METHODOLOGY

Introduction

The findings in this study were intended to influence policy and guidelines on teacher observation processes and feedback in a large urban school district. The findings examined the content of specific targeted feedback and addressed student knowledge gaps as measured by summative standardized assessments. Specific targeted feedback as defined by Rafalski (2015) is “The observer leaves differentiated and meaningful statements intended to improve the impact of an instructional strategy” (p. 79). Frequencies of level 1 and level 2 students connected to teachers who received specific targeted feedback were analyzed. In addition, correlations were calculated to examine the relationships between the categories and charges specific targeted feedback given as well as the elements coded in a teacher observation and student achievement level outcomes. The methodology used to interpret data is presented in this chapter through the following research questions:

1. What is the frequency of level 1 and 2 students in relationship to teachers who received specific targeted feedback?
2. What relationships, if any, exist between the specific targeted feedback as measured by elements scored during a school year and student achievement outcomes as measured by FCAT 2.0?

H₀₁. There are no significant relationships between the type of element scored during observations during a school year and student achievement outcomes as measured by FCAT 2.0.

3. What is the frequency by category of feedback, defined as content related feedback, pedagogy related feedback, procedural related feedback, coaching related feedback, or professional development related feedback provided by observers to teachers during classroom observations?
4. What relationships, if any, exist between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0?

H₀₂. There are no significant relationships between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0.

5. What relationships, if any, exist between the categories of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0?

H₀₃. There are no significant relationships between the categories of specific targeted feedback and student achievement outcomes as measured by FCAT 2.0.

For the purpose of this study, “teachers” were elementary, middle, and high school teachers who, as defined by Rafalski (2015), “...were instructional personnel in a large urban school district...had more than three years of teaching experience, and did not hold National Board Certification” (p. 68). Observations and feedback were provided through iObservation and the Marzano protocol. “Student achievement outcomes” were

acquired from the 2013-2014 FCAT 2.0 assessments. The FCAT 2.0 assessments used in this study tested reading, mathematics, and reading retake examinations. The subject areas and score ranges were dependent on the test and grade level in which the test was administered. Grades 3-10 were assessed in reading and grades 3-8 were assessed in FCAT 2.0 mathematics. Students in grades 11 and 12 who needed to retake the FCAT 2.0 Reading assessment took the grade 10 assessment.

Selection of Participants

The sample of this study consisted of 91 teachers. These teachers were elementary, middle, and high school teachers. They were all observed through the Marzano protocol and taught in the large urban school district in Central Florida. These teachers had more than three years of teaching experience and did not hold National Board Certification (Rafalski, 2015). In order to achieve a more in-depth analysis of feedback, the sample of 91 teachers was a purposive sample taken from a study conducted by Rafalski (2015); the original sample consisted of 2,718 teachers (p. 18). Only 91 teachers were reported as having received predominantly specific targeted feedback as measured by Rafalski's rubric (2015, p. 119). Table 1 shows the sample distribution of teachers who received different types of feedback, including specific targeted feedback. The focus of this study was the 91 teachers (3.35% of the original sample) who received specific targeted feedback from classroom observations (Rafalski, 2015).

Table 1
Teachers by Predominant Feedback Type

Predominant Feedback Type	<i>f</i>	% of Sample
Predominantly no feedback	282	10.38
Predominantly unrelated feedback	0	0.00
Predominantly recount of observation events	2,165	79.65
Predominantly general affirmations (praise)	104	3.83
Predominantly reflective feedback	53	1.95
Predominantly standardized feedback	23	0.85
Predominantly specific targeted feedback	91	3.35
Total Teachers	2,718	100.00

Source. Rafalski, S., H. (2015). Policy implications of a teacher evaluation system: The relationship of classroom observations, levels of feedback, and student achievement outcomes.

Instrumentation

The Marzano observation tool was used by observers in the large urban school district in Central Florida to collect observation data on the sampled teachers. “The psychometric adequacy for the teacher evaluation framework was supported by a research-based selection process. The Marzano Evaluation Model was chosen as the state model framework and was approved by the Florida Department of Education (DOE) for districts to use or adapt as their teacher evaluation model” (Rafalski, 2015, p. 69). The observation tool was accessed online through iObservation and provided teachers with electronic feedback on elements scored during observations.

Within this framework there were multiple domains that made up a teacher’s instructional practice score, which is the final quantitative measure of the elements coded and rated during classroom observations (LSI “Developing Policies and Procedures,” 2010). The Marzano learning map was segmented into four domains and further divided into 9 design questions (Marzano, “Teacher Evaluation Model,” 2011). Each domain

focused on a different aspect of teaching. The domains were as follows: “Domain 1: Classroom Strategies,” “Domain 2: Planning and Preparing,” “Domain 3: Reflecting on Teaching,” “Domain 4: Collegiality and Professionalism” (Marzano, “Teacher Evaluation Model,” 2011). During classroom observations, Domain 1 was the main focus as observers rated the instructional strategies utilized in the classroom at the time of the observation.

Within Domain 1 there were 41 elements that were “...revealed by research for effective teaching...” (Marzano, “Teacher Evaluation Model,” 2011). Elements 1 through 5 were “Involving Routine Events;” these elements focused on routines, rules, and procedures (Marzano, “Teacher Evaluation Model,” 2011). Elements 6 through 23 focused on “Addressing Content,” and honed in on the instruction of content (Marzano, “Teacher Evaluation Model,” 2011). These elements were ordered in a progression from teacher-led activities to student-centered activities. These elements were also organized into design questions 2, 3, and 4; as movement progressed down the learning map into design questions 3 and 4, students should have engaged in independent higher-level thinking (Marzano, “Teacher Evaluation Model,” 2011). Elements 24 through 41 were “Enacted on the Spot;” these elements were also procedural in nature and consisted of ways to engage students, adhere to rules and procedures, examine relationships teachers built with students, and view how teachers held high expectations for all students (Marzano, “Teacher Evaluation Model,” 2011). While all elements were important and contributed to student success, the large urban school district in Central Florida focused

more narrowly on elements 6 through 23 as they directly impacted student achievement through content delivery and cognitive processes.

Student performance data were measured through state standardized assessments. The assessment data examined were reading, mathematics, and reading retake achievement level outcomes from the 2013-2014 FCAT 2.0.

Proficiency levels were measured by specific score ranges as determined by the Florida Department of Education. The following figures depict the score ranges for each grade level and subject area assessed. It is important to remember that this study analyzed data from FCAT initial and retake assessments. Students who did not initially pass FCAT 2.0 Reading were scored on the same scale when reassessed as the test was from grade 10. For example, a student in grade 11 who needed to retake the FCAT 2.0 Reading assessment was scored on the grade 10 developmental scales. Figures 3 and 4 provide examples of achievement levels for the FCAT 2.0 Reading and FCAT 2.0 Mathematics assessments.

Table 1: Achievement Levels for the FCAT 2.0 Reading Developmental Scale Scores (140 to 302)

Grade	Level 1	Level 2	Level 3 (Passing)	Level 4	Level 5
3	140-181	182-197	198-209	210-226	227-260
4	154-191	192-207	208-220	221-237	238-269
5	161-199	200-215	216-229	230-245	246-277
6	167-206	207-221	222-236	237-251	252-283
7	171-212	213-227	228-242	243-257	258-289
8	175-217	218-234	235-248	249-263	264-296
9	178-221	222-239	240-252	253-267	268-302
10	188-227	228-244	245-255	256-270	271-302

Figure 4 Achievement levels for FCAT 2.0 Reading assessment grades 3 through 10
 Source: Florida Department of Education/Bureau of K-12 Student Assessment

**Table 2: Achievement Levels for the
FCAT 2.0 Mathematics Developmental Scale Scores (140 to 298)**

Grade	Level 1	Level 2	Level 3 (Passing)	Level 4	Level 5
3	140-182	183-197	198-213	214-228	229-260
4	155-196	197-209	210-223	224-239	240-271
5	163-204	205-219	220-233	234-246	247-279
6	170-212	213-226	227-239	240-252	253-284
7	179-219	220-233	234-247	248-260	261-292
8	187-228	229-240	241-255	256-267	268-298

Figure 5 Achievement levels for FCAT 2.0 Mathematics assessment grades 3-8
Source: Florida Department of Education/Bureau of K-12 Student Assessment

According to the 2014 Technical Report, the Florida Department of Education reported reliability and validity for the FCAT 2.0 were strong and properly scored; however, further studies may have been needed for extrapolation and implication purposes (p. 137).

Data Collection

De-identified data were taken from the large urban school district database. The data request consisted of the 91 teachers identified in Rafalski’s study (2015) as having received predominantly specific targeted feedback through iObservation and the students’ achievement level outcomes who were attached to those personnel numbers. While collecting data, this researcher learned that more than 91 teachers received feedback that was scored as specific targeted feedback according to Rafalski’s (2015) rubric. The focus of this study is only on the sample of teachers who received “predominantly” specific targeted feedback in their observations, which as defined by Rafalski (2015) is “Teachers...placed in the respective feedback categories based on the feedback type that represented that majority or plurality of the feedback received” (p. 82).

For the purpose of this study a breakdown of the content of specific targeted feedback was conducted to categorize what feedback was provided to teachers, and which feedback was most helpful in addressing student knowledge voids. To do this a rubric was created to give specific information on the categories of specific targeted feedback given to teachers. The rubric was created by using the sample rubric provided in appendix e of Rafalski's (2015) study and focused only on the column that related to specific targeted feedback. Rafalski's (2015) sample rubric can be found in appendix C of this study. The sample rubric was analyzed and content from the observation data helped determine how to group the specific targeted feedback for this study.

This rubric, which can be found in greater detail in appendix d, is as follows:

Content related: The feedback is related to the specific content or subject-area being taught.

Pedagogy related: The feedback is related to specific teaching practices or strategies.

Procedural: The feedback is related to classroom rules, procedures, and teacher rapport.

Coaching: The observer gives an explicit coaching tip or refers the teacher to other professionals for support.

Professional development: The observer suggests specific professional development to the teacher to improve instruction.

It is of importance to note that comments may have consisted of components for more than one category but were only coded once on their predominant quality. For example, a comment may have consisted of a coaching tip and a procedural note. The comment would have been coded based on what the feedback emphasized.

The charge of the specific targeted feedback was also examined to determine if positive, negative, or neutral feedback contributed to the relationship of feedback to assessment outcomes. A rubric for determining the charge of feedback was also created for this study and can be found in greater detail in appendix e.

The rubric for charge of feedback is as follows:

Neutral: Of or relating to possessing good or desirable qualities.

Negative: Of or relating to possessing qualities that are undesirable.

Positive: Having no strong good or undesirable qualities.

It is of importance to note that comments may have consisted of components for more than one charge of feedback but were only coded once on their predominant quality. For example, a comment may have consisted of positive feedback and neutral feedback. The comment would have been coded based on what the feedback emphasized.

This study was conducted using multiple pieces of archival data: teacher observation comments and historical student standardized assessment scores. This study utilized a mixed method to analyze these data. A qualitative method of data collection was used to help the researcher further delineate different categories of feedback and how they were related to student achievement outcomes or knowledge voids and a quantitative method was used to analyze the frequencies and relationships of specific targeted feedback and assessment scores.

Permission to use data was requested and granted from two entities: The Large Urban School District and The University of Central Florida. Written consent was

obtained through the Accountability, Research, and Grants Department from the school district in this study as well as from the University of Central Florida Institutional Review Board (IRB). The data request included a specific sample of “de-identified iObservation teacher protocol data including grade level designation, scale ratings and comments” (Rafalski, 2015, p. 77) and de-identified student FCAT 2.0 data. FCAT 2.0 data consisted of achievement levels 1-5 from reading, mathematics, and reading retake assessments. The de-identified teacher observation data was further narrowed to only include the 91 teachers who received predominantly specific targeted feedback as identified in Rafalski’s (2015) study.

Data Analysis

The research questions in this study were analyzed using descriptive statistics, frequencies, and other statistical tests such as Pearson’s r to study correlations between feedback and student achievement data. Descriptive statistics were used to gather information on mean, standard deviation, and frequencies of data. Frequencies were used to “...list each observed score, along with the number of cases falling at each score” (Steinberg, 2011, p. 23).

Correlations must be used with a single group of subjects (Steinberg, 2011). In this study, the single group of subjects was the 91 teachers who received predominantly specific targeted feedback as defined by Rafalski (2015). Correlations must have a score on two different variables (Steinberg, 2011). The scores used in this correlational study were the mean achievement levels from the multiple FCAT 2.0 assessment subject areas

and either a category or charge of feedback or a Marzano element. For example, correlations examined the relationship between feedback that contained coaching tips and the mean achievement levels of the FCAT 2.0 Reading assessment. The correlation determined if the scores on the two variables were related, not if there was a cause-and-effect relationship (Steinberg, 2011, p. 417). The purpose of using a correlational test for this study was to examine if a specific type of feedback or feedback delivered a certain way would predict student achievement. For example, the research questions sought to determine if teachers received more positive feedback would they have higher student achievement outcomes on FCAT 2.0?

Correlations have both a relationship strength and direction (Steinberg, 2011). The strength showed the degree to which the variables were related to one another and was measured on a scale from .00 to 1.00. The closer to 1.00 the correlation was, the stronger the relationship, no matter the sign (positive or negative) (Steinberg, 2011, p. 422). An example to illustrate relationship strength is as follows: .00 indicates no relationship, .12 is a weak relationship, and .80 is a strong relationship (Steinberg, 2011, p. 422). It is important to take note that in the social sciences, relationships are usually only moderately related because of other variables that affect human beings.

The direction of the relationship was also important to understand as it "...tells whether or not the values on two variables go up and down together" and "...is indicated by a positive or negative sign" (Steinberg, 2011, p. 423). A positive correlation meant as the scores on one variable increased the scores on the other variable increased as well

while a negative correlation meant as the scores on one variable increased the scores on the other variable decreased (Steinberg, 2011, p. 423).

The research questions in this study that analyzed correlations also had a null hypothesis that stated there was no correlation between the student achievement outcomes and specific feedback, which would produce a correlation of 0. There was no expectation that results would show a correlation of 0, so it was determined if the correlation was statistically significant (Steinberg, 2011, p. 439). If the correlation showed significance, then the null hypothesis was rejected. If the correlation was not significant then the null hypothesis was retained. To be considered a lot, the correlation coefficient must have matched or exceed the level of significance for a 2-tailed test (Steinberg, 2011).

Analyzing the statistical significance was only the beginning of examining the importance of the data. The practical importance was also considered, which determined the usefulness of the information (Steinberg, 2011). To determine if the correlation was small, medium, or large and its practical importance, the following effect size guidelines were used: small = .25 or less, medium = .25 to .40, and large = .40 or more (Steinberg, 2011, p. 447).

Pearson's r was used to determine relationships between variables. Relationships should not be confused with causation. Pearson's r measured relationships between scores, or "...the linear relationship between two variables that have both been measured on at least an interval level" (Steinberg, 2011, p. 432). The coefficient "...told the strength and the direction of a relationship...with -1.00 indicating a perfect negative

relationship, .00 indicating no relationship, and +1.00 indicating a perfect positive relationship” (Steinberg, 2011, p. 432). This correlation coefficient will tell the strength and direction of the relationship between the category of feedback, charge of feedback, or Marzano elements and the mean student achievement level outcomes for the 2013-2014 FCAT 2.0 Reading, Reading Retake, and Mathematics assessments. This test was most appropriate to use because the data used for each variable was a score.

Each research question was answered using the following statistical procedures:

Research Question 1

What is the frequency of level 1 and 2 students in relationship to teachers who received specific targeted feedback?

To answer Research Question 1 descriptive statistics were run to show the frequencies of level 1 and level 2 students. Data for all students who were attached to the 91 teachers who received predominantly specific targeted feedback were collected. Data were examined to see how many students took each assessment and the percentages of level 1 and level 2 students associated with each assessment. Comparative information for levels 3, 4, and 5 students was also included to answer research question 1.

Research Question 2

What relationships, if any, exist between the specific targeted feedback as measured by the elements scored during a school year and student achievement outcomes as measured by FCAT 2.0?

H₀₁. There are no significant relationships between the type of element scored during observations during a school year and student achievement outcomes as measured by FCAT 2.0.

This research question sought to examine what relationships, if any, existed between the frequencies of elements scored by observers during classroom observation and student achievement level outcomes as measured by FCAT 2.0. The desired outcome was to determine if specific elements coded during observations were related to improving student knowledge voids. Data were first sorted using frequencies to examine how many times each element was coded in this sample. The goal was to identify trends and gaps in instructional practice feedback. The researcher identified which elements were coded for each teacher and calculated the mean for each of the elements coded for each teacher. The researcher then calculated the mean achievement level outcome for each set of student data associated with the correct teacher. Finally, Pearson's r was used to calculate the correlations between the elements coded and student achievement level outcomes.

Research Question 3

What is the frequency by category of feedback, defined as content related feedback, pedagogy related feedback, procedural related feedback, coaching related feedback, or professional development related feedback provided by observers to teachers during classroom observations?

To answer Research Question 3, the determined what categories of specific targeted feedback observers provided to teachers and if the feedback was conducive to addressing student knowledge gaps as measured by FCAT 2.0. To achieve this, descriptive statistics were run and frequencies were calculated by the type of specific targeted feedback (content related feedback, pedagogy related feedback, coaching feedback, or professional development feedback) as recorded by observers. A rubric was developed to define the different types of specific targeted feedback. The rubric was developed based on a sub-rubric from Rafalski's (2015) study. The sub-rubric can be found in appendix c of this researcher's study. Key words were used to determine categories for the feedback. A new rubric was created with criteria for each category. The observation comments were read and categorized based on the rubric then later reread to check the coded categories. Descriptive statistics were then run to examine the frequencies of each category of feedback. The goal of this research question was to identify the distribution of the categorized feedback.

Research Question 4

What relationships, if any, exist between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT2.0?

H₀₂. There are no significant relationships between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0.

To answer Research Question number 4, descriptive statistics and frequency distributions were run and analyzed to determine how many comments given in the

specific targeted feedback were positive, negative, or neutral. The definition of positive, negative, and neutral feedback (charge) needed to first be determined. A rubric was then created with the criteria of what constituted feedback charge. Each feedback comment was read and coded based on the perceived charge. If more than one charge was present, the predominant charge was coded. Frequencies were run to determine the distribution of feedback charges from the sample. Once the comments were coded for each teacher, the average percentage of each charge was calculated for each teacher. Correlations were then calculated using Pearson's r to examine if there was a relationship between the charge of the feedback given during observations and student achievement outcomes as measured FCAT 2.0. The goal of this research question was to determine if a specific charge of feedback had a significant effect on student achievement outcomes.

Research Question 5

What relationships, if any, exist between the categories of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0?

H₀₃. There are no significant relationships between the categories of specific targeted feedback and student achievement outcomes as measured by FCAT 2.0.

Research question 5 examined the relationships between the categories of specific targeted feedback and student achievement level outcomes as measured by FCAT 2.0. The desired effect was to determine if a certain category of specific targeted feedback had a greater effect on addressing student knowledge voids of level 1 and 2 learners.

Data were first analyzed by examining the coded comments and frequency distributions from research question 3. After each comment was categorized, the feedback was averaged by category for each teacher. For example, if teacher 1 had 12 comments then each comment was calculated to determine the percentage for each category, so if teacher 1 had 3 coaching tips then that was calculated to 0.25, or 25 percent.

Once each percentage was calculated for each category of feedback, the mean achievement level for the students in each teacher's class was calculated. It is important to remember the data for this study consisted of elementary, middle, and high school teachers. Comparing grade level assessment outcomes was beyond the scope of this study, but knowing there were different levels is important to remember because the developmental assessment scores were measured differently for each grade level while the achievement levels were measured the same (1-5) for all grade levels. Therefore, the mean achievement level was a more accurate representation of data to be compared.

The mean achievement level score was calculated for each teacher and could range from 1-5 with 1 being the lowest and 5 being the highest. Proficiency began at level 3, and levels 1 and 2 were below proficient.

Once each mean achievement level was calculated, the data were then able to be correlated with the feedback categories. Pearson's r was used to calculate the relationships between the categories of feedback and mean student achievement level outcomes. Results were then examined for significance and practical importance.

Table 2 shows the research questions of the study. The variables, data sources, and methods of analysis are also presented in the table.

The next chapter analyzed the data in this study. Chapter 4 consists of in-depth analyses of the data for each research question. Data are discussed and illustrated with tables to support the calculated findings.

Table 2
Research Questions, Variables, Sources of Data and Methods of Analysis

Research Questions	Independent Variable(s)	Dependent Variable	Source(s) of Data	Method(s) of Analysis
1. What is the frequency of level 1 and 2 students in relationship to teachers who received specific targeted feedback?	<ul style="list-style-type: none"> • Teachers who received specific targeted feedback • Historical student assessment (FCAT 2.0) data from the 2013-2014 school year. 	Frequency of level 1 and 2 students	iObservation Reporting Student assessment data	Descriptive statistics and frequency distributions
2. What relationships if any exist between the specific targeted feedback as measured by elements scored during a school year and student achievement outcomes as measured by FCAT 2.0?	<ul style="list-style-type: none"> • Teachers who received specific targeted feedback. • Observation data for the 2013-2014 school year. • Historical student assessment (FCAT 2.0) data from the 2013-2014 school year. 	Student achievement outcomes	iObservation Reporting Historical student assessment data	Descriptive statistics Pearson's r
3. What is the frequency by content of feedback, defined as content related feedback, pedagogy related feedback, procedural related feedback, coaching related feedback, or professional development related feedback provided by observers to teachers during classroom observations?	<ul style="list-style-type: none"> • Observation data for the 2013-2014 school year. • Teachers who received targeted specific feedback 	Categories of specific targeted feedback	iObservation feedback	Descriptive statistics, frequency distributions, and qualitative coding

Research Questions	Independent Variable(s)	Dependent Variable	Source(s) of Data	Method(s) of Analysis
4. What relationships, if any, exist between the frequencies positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0?	<ul style="list-style-type: none"> • Observation data for the 2013-2014 school year. • Teachers who received targeted specific feedback • Historical student assessment (FCAT 2.0) data from the 2013-2014 school year. 	Types of specific targeted feedback Student assessment outcomes	iObservation feedback Historical student assessment data	Descriptive statistics, qualitative coding, frequency distributions, and Pearson's r.
5. What relationships, if any, exist between the categories of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0?	<ul style="list-style-type: none"> • Observation data for the 2013-2014 school year. • Teachers who received targeted specific feedback 	Student achievement outcomes	iObservation feedback Historical student assessment data	Pearson's r.

CHAPTER 4 ANALYSIS OF DATA

Introduction

The purpose of this study was to examine the content of specific targeted feedback as defined by Rafalski (2015) and to examine the relationships between feedback data and student achievement outcomes. The goal of this research was to discover where knowledge gaps existed in student achievement and determine what kind of specific targeted feedback teachers received in order to increase student achievement. 91 teachers received specific targeted feedback according to Rafalski's (2015) study, and the student assessment sample consisted of 18,875 total student scores. This includes students who took the spring 2013-2014 FCAT 2.0 in mathematics, reading, or reading retake assessments. Teacher observation and student assessment data were retrieved from the large urban school district for this study and analyzed through descriptive statistics and correlations. Results were reported through frequency tables and statistical tests and described narratively. These data analyses sought to answer the five research questions posed in this study. Chapter 4 contains the data, results, and reports from the statistical and qualitative analyses.

Research Question 1

What is the frequency of level 1 and level 2 students in relationship to teachers who received specific targeted feedback?

Data were provided by the large urban school district for this study. Data consisted of teacher observation data from the Marzano protocol from one school year (2013-2014). Student data consisted of mathematics, reading, and reading retake FCAT 2.0 assessment scores from the 2013-2014 school year. The teacher observation data from Rafalski’s (2015) study (2015) were narrowed to 91 teachers out of 2,718 teachers who were scored as receiving predominantly specific targeted feedback. The teacher requirements of this study “...were instructional personnel in a large urban school district...had more than three years of teaching experience, and did not hold National Board Certification” (Rafalski, 2015, p. 68). Tables 3, 4, 5, and 6 display the distribution of how many level 1 and level 2 students, as defined by the FCAT 2.0 assessment, were tied to the 91 teachers who received specific targeted feedback. Table 3 shows the student sample for each assessment.

Table 3
Description of Student Achievement Levels on FCAT 2.0 Assessments

Assessment	N	Missing	Mean
FCAT 2.0 Reading Achievement Level	9775	970	2.44
FCAT 2.0 Reading Retake Achievement Level	347	10398	2.12
FCAT 2.0 Mathematics Achievement Level	8753	1992	2.31
Total	18875		

This study consisted of achievement level data from 10,745 students. Many of these students had more than one assessment score, which resulted in a sample of 18,875 total assessment scores. The FCAT 2.0 Reading assessment had 9,775 student scores, which shows there were no scores for 970 students. To score proficiently on the FCAT

2.0 assessment, the achievement level must be a level 3, level 4, or level 5. The mean achievement level for all FCAT 2.0 Reading assessment scores in this sample was 2.44, which is below proficient. The FCAT 2.0 Reading Retake assessment had 347 student scores, which shows there were no scores for 10,398 students as the retake assessment is only for students who did not previously take the FCAT 2.0 Reading assessment or who did not pass the assessment the previous year. The mean achievement level for the FCAT 2.0 Reading Retake assessment was 2.12, which was below proficient. The FCAT 2.0 Mathematics assessment had 8,753 student scores, which shows there were no scores for 1,992 students. The mean achievement level for all FCAT 2.0 Mathematics assessments was 2.31, which is below proficient. Information based on these data showed all mean achievement levels for the three FCAT 2.0 assessments of this study were below proficient even though teachers received specific targeted feedback.

*Table 4
Frequency of Student Achievement Levels of Teachers with Specific Targeted Feedback on Reading FCAT 2.0*

Achievement Level	<i>f</i>	% of Students Assessed	% of Total Student Sample
Level 1	2,391	24.5	22.3
Level 2	3,116	31.9	29.0
Level 3	2,395	24.5	22.3
Level 4	1,336	13.6	12.4
Level 5	537	5.5	5.0
No Score	970	9.9	9.0
Total	10,745		100.0

According to the data, 9,775 (91.0%) students took the FCAT 2.0 reading assessment. There were no reading scores for 970 (9.9%) students included in the sample. While the focus of the research question was how many students scored at a level 1 or level 2, comparing those numbers to students who scored proficiently was

important to see the proportion of proficiency to deficiency. The sample consisted of 2,391 (24.5%) students who scored at a level 1 on the assessment and 3,116 (31.9%) students who scored at a level 2 on the FCAT 2.0 assessment. The total number of students who scored at level 1 or level 2 on the FCAT 2.0 Reading assessment was 5,507 or 56.3% of all students who took the reading assessment. Those who scored at level 3, level 4, or level 5 (proficient) consisted of 4,268 students or 43.7% of students who took the reading assessment. There were more students who scored at a level 1 or level 2 than students who scored proficiently on the FCAT 2.0 Reading assessment.

Table 4 shows the frequency distribution of students who took the FCAT 2.0 Mathematics assessment. Achievement levels were categorized into five levels with level 1 being the lowest and level 5 being the highest. To be proficient, students must score at least a level 3 on any FCAT 2.0 assessment.

Table 5
Frequency of Student Achievement Levels of Teachers with Specific Targeted Feedback on Mathematics FCAT 2.0

Achievement Level	<i>f</i>	% of Students Assessed	% of Total Student Sample
Level 1	2,638	30.1	24.6
Level 2	2,544	29.1	23.7
Level 3	2,172	24.8	20.2
Level 4	1,009	11.5	9.4
Level 5	390	4.5	3.6
No Score	1,992	22.8	18.5
Total	10,745		100.0

According to the data, 8,753 (81.5%) students took the FCAT 2.0 Mathematics assessment. There are no mathematics scores for 1,992 (22.8%) students. While the focus of the research question was on how many students scored at level 1 or level 2, comparing those numbers to students who scored proficiently was important to see the

proportion of proficiency to deficiency. The sample consisted of 2,638 (30.1%) students who scored at a level 1 on the FCAT 2.0 Mathematics assessment and 2,544 (29.1%) students who scored at a level 2 on the FCAT 2.0 assessment. The total number of students who scored at level 1 or level 2 on the FCAT 2.0 Mathematics assessment was 5,182 (59.2%) of all students who took the mathematics assessment. Those who scored at level 3, level 4, or level 5 (proficient) consisted of 3,571 students or 40.8% of students who took the mathematics assessment. There were more students who scored at a level 1 or level 2 on the FCAT 2.0 Mathematics assessment than students who scored proficiently.

Table 5 shows the frequency distribution of students who took the FCAT 2.0 Reading Retake assessment. Achievement levels were categorized into five levels with level 1 assigned the lowest level and level 5 assigned the highest level. To be proficient, students must have scored at least a level 3 on any FCAT 2.0 assessment.

Table 6
Frequency of Student Achievement Levels of Teachers with Specific Targeted Feedback on Reading FCAT 2.0 Retake

Achievement Level	<i>f</i>	% of Students Assessed	% of Total Student Sample
Level 1	89	25.6	0.8
Level 2	147	42.4	1.4
Level 3	93	26.8	0.9
Level 4	18	5.2	0.2
Level 5	0	0	0.0
Total Assessed	347		3.3

According to the data, 347 (3.2%) students took the FCAT 2.0 Reading retake assessment. There were no mathematics retake scores in this sample. While the focus of the research question was on how many students scored at level 1 or level 2, comparing

those numbers to students who scored proficiently was important to see the proportion of proficiency to deficiency. The sample consisted of 89 (25.6%) students who scored at a level 1 on the assessment and 147 (42.4%) students who scored at a level 2 on the assessment. The total number of students who scored at level 1 or level 2 on the FCAT 2.0 Reading retake assessment was 236 or 68.0% of all students who took the reading retake assessment. Those who scored at levels 3 or 4 (proficient) consisted of 111 students or 32.0% of students who took the reading retake assessment. There were no students who scored a level 5 on the FCAT 2.0 Reading retake assessment in this sample. There were more students who scored below proficient on the FCAT 2.0 Reading Retake assessment than students who scored at or above proficient levels.

Data from these results indicated that even though teachers received specific targeted feedback, their students scored predominantly at level 1 or level 2 on the FCAT 2.0 Reading assessment, Mathematics assessment, and Reading Retake assessment.

Research Question 2

What relationships, if any, exist between the specific targeted feedback as measured by elements scored during a school year and student achievement outcomes as measured by FCAT 2.0?

H₀₁. There are no significant relationships between the type of element scored during observations during a school year and student achievement outcomes as measured by FCAT 2.0.

This question sought to answer if there was a relationship between the specific elements scored as receiving specific targeted feedback and student assessment scores as measured by FCAT 2.0. The purpose for this question was to determine if outcomes from assessment scores was higher in relationship with feedback on certain Marzano elements. For example, did feedback on identifying critical content show a more positive relationship than feedback on using homework? Table 8 showed the frequency of each element that was scored as receiving specific targeted feedback. To determine if there was a relationship between the type of element coded and student achievement outcomes, Pearson's r was calculated to show correlations. Table 8 showed the significance of the relationships between the Marzano elements coded and mean student achievement level outcomes.

The data in table 7 showed the frequency of each element scored for the teachers who received specific targeted feedback. These data sought to answer the question of why students scored at level 1 and level 2 on assessments if their teachers predominantly received specific targeted feedback.

Table 7
Frequency of Element and Design Questions Scored Receiving Specific Targeted Feedback

Elements Coded (DQ = Design Question)	<i>f</i>	Percent
What will I do to establish and communicate learning goals, track student progress, and celebrate success? (DQ 1)		
1. Providing Clear Learning Goals and Scales (Rubrics)	65	14.4
2. Tracking Student Progress	15	3.3
3. Celebrating Success	5	1.1
What will I do to establish or maintain classroom rules and procedures? (DQ 6)		
4. Establishing Classroom Routines	32	7.1
5. Organizing the Physical Layout of the Classroom	12	2.7
What will I do to help students effectively interact with new knowledge? (DQ 2)		
6. Identifying Critical Content	9	2.0
7. Organizing Students to Interact with New Knowledge	18	4.0

Elements Coded (DQ = Design Question)	<i>f</i>	Percent
8. Previewing New Content	11	2.4
9. Chunking Content into “Digestible Bites”	21	4.7
10. Processing New Information	7	1.6
11. Elaborating on New Information	12	2.7
12. Recording and Representing Knowledge	14	3.1
13. Reflecting on Learning	4	0.9
What will I do to help students practice and deepen their understanding of new knowledge? (DQ 3)		
14. Reviewing Content	33	7.3
15. Organizing Students to Practice and Deepen Knowledge	17	3.8
16. Using Homework	3	0.7
17. Examining Similarities and Differences	5	1.1
18. Examining Errors in Reasoning	6	1.3
19. Practicing Skills, Strategies, and Processes	25	5.5
20. Revising Knowledge	2	0.4
What will I do to help students generate and test hypothesis about new knowledge? (DQ 4)		
21. Organizing Students for Cognitively Complex Tasks	-	-
22. Engaging Students in Cognitively Complex Tasks Involving Hypothesis Generation and Testing	1	0.2
23. Providing Resources and Guidance	-	-
What will I do to engage students? (DQ 5)		
24. Noticing When Students Are Not Engaged	40	8.9
25. Using Academic Games	3	0.7
26. Managing Response Rates	30	6.7
27. Using Physical Movement	-	-
28. Maintaining a Lively Pace	19	4.2
29. Demonstrating Intensity and Enthusiasm	2	0.4
30. Using Friendly Controversy	1	0.2
31. Providing Opportunities for Students to Talk About Themselves	1	0.2
32. Presenting Unusual or Intriguing Information	1	0.2
What will I do to recognize and acknowledge adherence and lack of adherence to classroom rules and procedures? (DQ 7)		
33. Demonstrating “Withitness”	12	2.7
34. Applying Consequences for Lack of Adherence to Rules and Procedures	16	3.5
35. Acknowledging Adherence to Rules and Procedures	2	0.4
What will I do to establish and maintain effective relationships with students? (DQ 8)		
36. Understanding Students’ Interests and Background	1	0.2
37. Using Verbal/Nonverbal Behaviors that Indicate Affection for Students	1	0.2
38. Displaying Objectivity and Control	2	0.4
What will I do to communicate high expectations for all students? (DQ 9)		
39. Demonstrating Value and Respect for Low Expectancy Students	-	-
40. Asking Questions of Low Expectancy Students	1	0.2
41. Probing Incorrect Answers with Low Expectancy Students	2	0.4
Total	451	100.0

Note. Dashes indicate areas where data could not be reported because data were not obtained.

Data from this table shows the frequency of elements coded as having received specific targeted feedback. This is important as it was clear which elements administrators focused on when giving feedback to teachers. The data also highlighted which elements were not scored as often or even neglected in terms of receiving specific targeted feedback. Data indicated that the element scored most often was from design question 1, element 1: Providing Clear Learning Goals and Scales. Learning goals and scales give students information about what they will be learning and provides an avenue for students to track their progress as they advance through a lesson or unit (Marzano, 2012). The areas that received the lowest amount of specific targeted feedback were from design questions 4, 8, and 9. Design question 4 (What will I do to help students generate and test hypothesis about new knowledge?) was coded only one time (0.2% of feedback). Design question 8 (What will I do to establish and maintain effective relationships with students?) was coded only four times (0.9% of feedback). Design question 9 (What will I do to communicate high expectations for all students?) was coded only three times (0.7% of feedback). Four elements received no specific targeted feedback. Two of these elements were from design question 4: element 21-organizing students for cognitively complex tasks and element 23-providing resources and guidance. One element is from design question 5: element 27-using physical movement. One element was from design question 9: element 39- demonstrating value and respect for low expectancy students. Design question 4 was of importance because of the level of thinking that is required of students, and design questions 8 and 9 were important because of the implications they have on students.

Table 8 showed the correlations between the elements scored during teacher observations and student mean student achievement levels for each FCAT 2.0 assessment. Correlations existed for almost all elements to some degree. Negative correlations were represented with a negative sign in front of the decimal, and positive correlations began with a decimal. Elements that had significant correlations were marked with either an asterisk or a double asterisk depending on the confidence level of the correlation.

Table 8
Pearson Correlations between Elements Scored and FCAT 2.0 Mean Student Achievement Levels

Element	Mean Reading Achievement Level	Mean Reading Retake Achievement Level	Mean Mathematics Achievement Level
1. Providing clear learning goals and scales	-.005	-.152	-.013
2. Tracking student progress	-.097	-.068	.005
3. Celebrating success	.244*	-.065	.208*
4. Establishing classroom rules and procedures	-.172	.152	-.123
5. Organizing the physical layout of the classroom	-.150	-.065	-.068
6. Identifying critical content	.013	-.089	.078
7. Organizing students to interact with new knowledge	.072	-.011	.076
8. Previewing new content	.116	.170	-.177
9. Chunking content into “Digestible bites”	-.018	.112	-.003
10. Processing of new information	-.009	.229*	-.058
11. Elaborating on new information	-.117	.073	-.090
12. Recording and representing knowledge	.012	.128	.000
13. Reflecting on learning	.120	-.050	.087
14. Reviewing content	.028	-.060	.078
15. Organizing students to	-.099	-.014	-.049

Element	Mean Reading Achievement Level	Mean Reading Retake Achievement Level	Mean Mathematics Achievement Level
practice and deepen knowledge			
16. Using homework	.203	-.049	.160
17. Examining similarities and differences	.101	-.055	.190
18. Examining errors in reasoning	.176	-.064	.201
19. Practicing skills, strategies, and processes	.056	.009	.078
20. Revising knowledge	.130	.054	.087
21. Organizing students for cognitively complex tasks	-	-	-
22. Engaging students in cognitively complex tasks involving hypothesis generation and testing	.005	-.030	.004
23. Providing resources and guidance	-	-	-
24. Noticing when students are not engaged	.028	-.102	.064
25. Using academic games	.052	-.052	.044
26. Managing response rates	-.084	-.016	-.068
27. Using physical movement	-	-	-
28. Maintaining a lively pace	.095	-.086	.110
29. Demonstrating intensity and enthusiasm	.045	-.042	.103
30. Using friendly controversy	.131	-.030	.107
31. Providing opportunities for students to talk about themselves	-.036	-.030	-.026
32. Presenting unusual or intriguing information	-.062	-.030	-.022
33. Demonstrating “withitness”	-.109	-.085	-.126
34. Applying consequences for lack of adherence to rules and procedures	-.141	.515**	-.330**
35. Acknowledging	-.001	-.043	.068

Element	Mean Reading Achievement Level	Mean Reading Retake Achievement Level	Mean Mathematics Achievement Level
adherence to rules and procedures			
36. Understanding students' interest and background	-.013	-.030	.011
37. Using verbal and nonverbal behaviors that indicate affection for students	.037	-.030	-.207*
38. Displaying objectivity and control	-.050	.295**	-.209*
39. Demonstrating value and respect for low expectancy students	-	-	-
40. Asking questions of low expectancy students	.092	-.030	.067
41. Probing incorrect answers with low expectancy students	-.006	-.041	.015

** Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed)

Note. Dashes indicate areas where data could not be reported because data were not obtained.

Table 8 showed the relationship between the elements coded during observations and student achievement level outcomes for FCAT 2.0 Reading, FCAT 2.0 Reading Retake, and FCAT 2.0 Mathematics assessments. Pearson's r was used to calculate correlations. Correlations as measured by Steinberg (2011) were defined as small = .25 or less, medium = .25 to .40, or large = .40 or more (p. 447). Correlations were also reported as a positive or negative direction. A positive correlation existed if the dependent variable increased as the independent variable increased, a negative correlation existed if the dependent variable decreased as the independent variable increased. Data from the correlations revealed only five elements had any statistically significant

relationships. Elements 3, 10, 34, 37, and 38 had either positive or negative correlations between elements coded and mean student achievement level outcomes as measured by FCAT 2.0.

Element 3 (celebrating success) in design question 1 had a significant correlation at the .05 level in relationship to mean reading achievement levels and a frequency of 5 comments (1.1%). The correlation, .244 was a positive correlation; however, it was a small correlation. The relationship suggested that as feedback in element 3 increased, student achievement increased on the FCAT 2.0 reading assessment. The significance at the .05 level indicated there was indeed a correlation with 95% confidence. So, $r(89) = .244, p < .05$. Pearson r at 89 degrees of freedom was .244. There was less than a 5% chance that this correlation was due to mere chance.

Element 3, celebrating success, also showed a correlation with the mean mathematics achievement level at .208. This correlation was a small and positive correlation, which indicated a prediction that as feedback in element 3 increased, student achievement on the FCAT 2.0 Mathematics assessment increased. The correlation was significant at the .05 level indicating there was indeed a correlation with 95% confidence. So, $r(89) = .208, p < .05$. Pearson r at 89 degrees of freedom was .208. There was less than a 5% chance that this correlation was due to mere chance.

In terms of practical importance, a .244 correlation and a .208 correlation were both too small to recommend administrators focus solely on element 3 when conducting classroom observations. However, since element 3 did have a correlation with statistical significance, it may be prudent to include element 3 in conjunction with other elements

during observations. It is also of importance to note since the frequency of this element had only 5 comments the validity of the correlation may be in question.

Element 10, processing of new information, was another element that had a significant relationship with the mean reading retake achievement level outcomes and a frequency of 7 comments (1.6%). The correlation between these two variables was .229, a small and positive correlation. This correlation indicated a prediction that as feedback in element 10 increased, student achievement levels for the FCAT 2.0 Reading Retake assessment increased. The correlation was significant at the .05 level, which indicated a correlation with 95% confidence. Therefore, $r(89) = .229, p < .05$. Pearson r at 89 degrees of freedom was .229. There was less than a 5% chance that this correlation was due to mere chance.

In terms of practical importance, a .229 correlation was too small to recommend administrators focus solely on element 10 when conducting classroom observations. However, since element 10 did have a correlation with statistical significance, it may be prudent to include element 10 in conjunction with other elements during observations. It is also of importance to note since the frequency of this element had only 7 comments the validity of the correlation may be in question.

Element 34, applying consequences for lack of adherence to rules and procedures, was calculated as having significant correlations with the mean reading retake achievement level and mean mathematics achievement level variables. This element also had a frequency of 16 comments (3.5%). The correlation between element 34 and the mean reading retake achievement levels was .515, which was a medium, positive

correlation. This indicated a prediction that as feedback in element 34 increased, mean student achievement levels increased on the FCAT 2.0 Reading Retake assessment. The correlation was significant at the .01 level, which indicated a correlation with 99% confidence. Therefore, $r(89) = .515$, $p < .01$. Pearson r at 89 degrees of freedom was .515. There was less than a 1% chance that this correlation was due to mere chance. In terms of practical importance, a .515 correlation is moderate, and including feedback in element 34 could be helpful for teachers with students taking reading retake assessments. It is also of importance to note since the frequency of this element had only 16 comments the validity of the correlation may be in question, yet this element had the highest frequency of comments out of the other 4 elements with statistically significant relationships.

Another correlation in element 34 was with the mean mathematics achievement levels. The correlation calculated at $-.330$, which was a medium, negative correlation. These data indicated that as feedback in element 34 increased, student achievement levels on the FCAT 2.0 Mathematics assessment decreased. The correlation was significant at the .01 level, which showed 99% confidence, that there was indeed a correlation and not a sampling error. So, $r(89) = -.303$, $p < .01$. Pearson r at 89 degrees of freedom was $-.303$. There was less than a 1% chance that this correlation was due to mere chance. In terms of practical importance, a $-.303$ correlation was moderate and had a negative impact on the mean mathematics achievement levels. Based on these data, element 34 may not be helpful to increasing student achievement in the area of mathematics. It is

also of importance to note since the frequency of this element had only 16 comments the validity of the correlation may be in question.

Applying consequences for lack of adherence to rules and procedures speaks to classroom management. The strength of the correlation in relationship to mean reading retake levels suggested using this element could have a positive effect on student achievement; however, it is of interest to determine why there would be a negative correlation in the area of mathematics.

Element 37, using verbal and nonverbal behaviors that indicated affection for students, showed a significant correlation with the mean mathematics achievement level variable and a frequency of 1 comment (0.2%). The correlation was $-.207$, which was a small and negative correlation. These data indicated a prediction that as feedback in element 37 increased, student achievement levels on the FCAT 2.0 Mathematics decreased. This correlation was significant at the 0.05 level, which showed 95% confidence the correlation was indeed significant and not due to sampling error. Therefore, $r(89) = -.207, p < .05$. Pearson r at 89 degrees of freedom was $-.207$. There was less than a 5% chance that this correlation was due to mere chance. Based on these data, it may not be helpful for administrators to incorporate isolated feedback in this element into observations of teachers who have students taking mathematics assessments. It is also of importance to note since the frequency of this element had only 1 comment the validity of the correlation may be in question.

Element 38, displaying objectivity and control, was the last element that had any significant correlations in this study and frequency of 2 comments (0.4%). There were

correlations with the mean reading retake achievement levels and with the mean mathematics achievement levels. The correlation with the mean reading retake achievement levels was .295, which was a medium, positive correlation. These data indicated a prediction that as feedback in element 38 increased, student achievement levels on the FCAT 2.0 reading retake assessment increased. This correlation was significant at the .01 level, which showed a 99% confidence level that the correlation was not due to a sampling error. Therefore, $r(89) = .295, p < .01$. Pearson r at 89 degrees of freedom was .295. There was less than a 1% chance that this correlation was due to mere chance. In terms of practical importance, a .295 correlation was moderate and had a positive impact on reading retake student achievement levels. Based on these data, it may be helpful for administrators to incorporate feedback in this element into observations of teachers who have students taking reading retake assessments. It is also of importance to note since the frequency of this element had only 2 comments the validity of the correlation may be in question.

The correlation between element 38 and the mean mathematics achievement levels was -.209. This was a small and negative correlation, which indicated a prediction that as feedback in element 38 increased student achievement levels on the FCAT 2.0 Mathematics assessment decreased. This correlation was significant at the .05 level, which showed a confidence level of 95% that the correlation was not due to a sampling error. So, $r(89) = -.209, p < .05$. Pearson r at 89 degrees of freedom was -.209. There was less than a 5% chance that this correlation was due to mere chance. In terms of practical importance, a -.209 correlation was small and had a negative impact on

mathematics achievement levels. Based on these data, it may not be helpful for administrators to incorporate isolated feedback in this element into observations of teachers who have students taking mathematics assessments. It is also of importance to note since the frequency of this element had only 2 comments the validity of the correlation may be in question.

While there were 5 elements that showed significant correlations with certain FCAT 2.0 assessments, it is important to take into considerations the sample size used. Although the correlations were predictions on what outcomes could be, the fact that only a small amount of comments were coded for these 5 elements, as well as the fact that very few correlations were significant should be factored into any potential decisions made about focusing on specific elements during observations.

Research Question 3

What is the frequency by category of feedback, defined as content related feedback, pedagogy related feedback, procedural related feedback, coaching related feedback, or professional development related feedback provided by observers to teachers during classroom observations?

To answer this question, the raw data from teacher observation comments were coded into categories. The data set provided by Large Urban School District consisted of 91 teachers' observation comments. These comments were first coded by Rafalski (2015) to examine if teachers predominantly received specific targeted feedback. Rafalski (2015) created a sub-rubric located in appendix c that gave stems and examples

of what constituted specific targeted feedback. To further code the specific targeted feedback, the sub-rubric was used as a starting point to identify key words. A rubric was then created to classify the specific targeted feedback into categories: coaching tips, procedural, pedagogy, professional development, and content. The raw data were then searched using key words and analyzed to determine which category of specific targeted feedback the comment belonged. Each comment was read and coded and then checked again with the rubric. The feedback were finally coded and sorted into specific categories as defined by the rubric.

Table 9 showed the distribution of the 451 specific targeted feedback comments separated into categories of content related feedback, pedagogy related feedback, coaching feedback, professional development related feedback, and procedural related feedback. Content related feedback was defined as feedback related to the specific content or subject-area being taught. Pedagogy related feedback was defined as feedback that was related to specific teaching practices or strategies. Coaching related feedback was defined as the observer gave an explicit coaching tip or referred the teacher to other professionals for support. Professional development related feedback was defined as feedback where the observer suggested specific professional development to the teacher to improve instruction. Procedural feedback was defined as feedback related to classroom rules, procedures, and teacher rapport. It is of importance to note that comments may have consisted of components for more than one category or charge of feedback but were only coded once on their predominant quality.

Table 9
Frequency of Feedback Content Coded

Code	<i>f</i>	Percent
Coaching Tip	266	59.0
Procedural	91	20.2
Pedagogy	59	13.1
Professional Development	28	6.2
Content	7	1.6
Total	451	100.0

Descriptive statistics showed the frequency and distribution of each category of specific targeted feedback. Data showed the most frequent type of feedback was coaching tips. 266 (59.0%) comments consisted of a coaching tip for the teacher to change instruction or incorporate into the classroom. Data were not coded based on the quality of the coaching tip, only if a coaching tip was given. 91 (20.2%) comments were coded as procedural. This type of feedback consisted of comments that referenced behavior, classroom rules and procedures, and layout and organization of the classroom. 59 comments (13.1%) were coded as pedagogy. Pedagogy and coaching tips were closely related but delineated based on whether the observer gave a suggestion/recommendation or commented on the teacher monitoring or making a statement reflective of pedagogy. 28 (6.2%) comments were coded as professional development. Professional development consisted of comments that referred teachers to the Marzano literature, referred teachers to other instructors, or made suggestions to videotape or record a lesson. Again the quality and validity of the professional development was not evaluated, the purpose was to identify if observers were directing teachers toward resources to help improve instruction. The smallest category of specific targeted feedback was content related. There were 7 (1.6%) comments related to content.

Content related comments were identified by feedback specifically related to a subject area. Content related feedback was not coded based on the accuracy of the feedback in relationship to the content, but if the observer made a content connection within the feedback.

Research Question 4

What relationships, if any, exist between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT2.0?

H₀₂. There are no significant relationships between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0.

This question was answered through coding of specific targeted feedback and descriptive statistics. A rubric was created (appendix E) to determine the charge of the specific targeted feedback. Data were coded based on if the feedback was predominantly negative, neutral, or positive. Each comment was read and coded by predominant charge of the feedback. Table 9 shows the distribution of the charge of specific targeted feedback: neutral, negative, and positive.

Data were coded based on key words in the observation comments. If comments contained praise or positive recognition, they were coded as positive. If comments focused on what teachers did incorrectly or what the teachers should not do, they were coded as negative. If comments were suggestions or recommendations that neither praised nor criticized teachers, they were coded as neutral. Pearson's r was calculated to

determine if a relationship existed. This question related back to Bandura’s social cognitive theory: If a person feels they are unable to do something then their behaviors will not change, while if a person is made to feel they are able to be successful their behaviors may reflect the desired outcome (Bandura, 1977). The prediction was that positive feedback would result in more behavior changes than negative or neutral feedback.

Table 10 revealed the relationship between the charge of the comments and student assessment scores on FCAT 2.0. There were 451 comments that were coded into the specific feedback charge categories and three charges of specific targeted feedback: Neutral, negative, and positive. Neutral feedback was defined as having no strong good or undesirable qualities. Negative feedback was defined as of or relating to qualities that are undesirable. Positive feedback was defined as of or relating to good or desirable qualities

*Table 10
Frequency of Feedback Charge*

Charge	<i>f</i>	Percent
Neutral	227	50.3
Negative	147	32.6
Positive	77	17.1
Total	451	100.0

The majority of feedback was coded as neutral feedback. 227 (50.3%) comments were coded as neutral. Neutral feedback was determined by the lack of praise or criticism in the feedback. Most neutral statements included what was observed and a categorical comment as defined by the content rubric, but did not include emotion-inciting comments. The tone was mostly matter-of-fact and while it could have included

something the teacher did well or needed to improve; it was not strongly worded. For example, the observer may suggest the students work together because collaboration increases rigor or ask the teacher to consider strategies to encourage all students to participate during a lesson. These examples would be considered neutral because there is no strong emphasis on good or undesirable qualities.

The second highest frequency of comment charge distribution was negative comments. 147 (32.6%) comments were coded as negative. Negative comments were coded based on the predominant tone of the feedback consisting of criticism of the teacher. Most negative feedback discussed what the teacher did wrong, how a strategy did not work, and how students were not on task, engaged, or compliant. For example, the observer asked students to explain what their learning goal was and they were all unable to explain it, or there was no learning goal posted, or the group sizes were too large or too small, or if the teacher did not redirect students or needed to redirect students too much. These examples would be considered negative feedback because they focused on the undesirable qualities.

The lowest frequency distribution of comment charge distribution was positive feedback. 77 (17.1%) comments were coded as positive. Positive feedback consisted of comments related to praise. The praise could be subtle or direct; some praise told teachers they were making great improvements or doing a good job; and some praise recognized something good going on in the classroom but still consisted of ways the teacher could improve. For example, an observer may have said great job/strategy and to be sure to continue to use a strategy and monitor all students. For example, the observer

noted that students were able to stay on task with the teacher during transitions and needed little to no redirection, the teacher used a great strategy that kept students interested and engaged, or the teacher created a classroom environment that supported student learning. These examples were considered positive because they emphasized the good or desired qualities in the lesson. While praise alone is not recommended when giving feedback (Hattie, 2009), the nature of the specific targeted feedback eliminated the possibility of a comment being empty praise by including items such as coaching tips to improve the lesson, or how the teacher could move to the next level on the observation scale.

To determine if there were relationships between the charge of feedback and FCAT 2.0 Reading, Reading Retake, and Mathematics achievement level outcomes Pearson's r was calculated. Table 11 illustrated the results of Pearson's r between the charge of the feedback and student achievement level outcomes.

*Table 11
Pearson Correlations between Charge Code and FCAT 2.0 Mean Student Achievement Levels*

Charge Code	Mean Reading Achievement Level	Mean Reading Retake Achievement Level	Mean Mathematics Achievement Level
Neutral	-.166	-.006	-.015
Negative	.067	.002	-.095
Positive	.118	.004	.112

Based on these data, there appeared to be very weak correlations and no statistically significant relationships between feedback charge and student achievement level outcomes. It may be of importance to note that although the correlations were not calculated as significant, the positive feedback had the highest correlations to student achievement level outcomes for all three FCAT 2.0 assessments.

Research Question 5

What relationships, if any, exist between the categories of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0?

H₀₃. There are no significant relationships between the categories of specific targeted feedback and student achievement outcomes as measured by FCAT 2.0.

To answer this research question, data were first coded by a rubric (appendix D) to determine if feedback was content related feedback, pedagogy related feedback, coaching feedback, or professional development feedback. Once each teacher's comments were coded, percentages for each code were calculated. The percentages were then measured in relationship to mean student achievement level outcomes. Pearson's r was used to determine if correlations existed between the type of feedback given to teachers and student achievement outcomes on FCAT 2.0. Table 11 shows the relationship between the content of the specific targeted feedback and student achievement as measured by assessment scores on FCAT 2.0.

Table 12
Pearson Correlations between Comment Code and FCAT 2.0 Mean Student Achievement Levels

Comment Code	Mean Reading Achievement Level	Mean Reading Retake Achievement Level	Mean Mathematics Achievement Level
Coaching Tip	.032	-.030	.037
Content	.054	-.041	.047
Procedural	-.183	.243*	-.247*
Professional Development	.045	.012	.039
Pedagogy	.076	-.185	.137

*Correlation is significant at the 0.05 level (2-tailed).

These data show the relationship between the type of specific targeted feedback each teacher received and the mean student achievement level of FCAT 2.0 Reading, Reading Retake, and Mathematics assessments. Pearson's r was calculated to determine if there were statistically significant relationships between the type of feedback given during observations and student achievement level outcomes as measured by FCAT 2.0. Data indicated there were no statistically significant relationships between the content of specific targeted feedback and student achievement scores in all categories except for one: Procedural related feedback.

Feedback related to coaching tips, content, professional development, and pedagogy showed no statistically significant relationships to student achievement outcomes. The procedural category showed statistically significant relationships between the feedback and student achievement outcomes on the FCAT 2.0 Reading Retake assessment (.243) and the FCAT 2.0 Mathematics assessment (-.247). Procedural feedback had a frequency of 91 comments (20.2%) out of 451 comments.

The relationship between procedural feedback and FCAT 2.0 Reading Retake assessment is small and positive which indicated a prediction that as procedural feedback was given, reading retake achievement level outcomes increased. Therefore, $r(89) = +.243$, $p < .05$. Pearson's r at 89 degrees of freedom was $+.243$. There was less than a 5% chance that this correlation was due to mere chance.

To determine the importance of the correlation, the strength and direction of the correlation were considered. The correlation was positive, which was a desired effect but overall it was relatively small. This showed that while procedures had a significant

relationship with student achievement level outcomes on the FCAT 2.0 Reading Retake assessment, it was not a strong relationship.

The relationship between procedural feedback and the FCAT 2.0 Mathematics assessment was a small, negative correlation, which indicated a prediction that as procedural feedback was given, FCAT 2.0 Mathematics assessment scores decreased. Therefore, $r(89) = -.247, p < .05$. Pearson r at 89 degrees of freedom was $-.247$. There was less than a 5% chance that this correlation was due to mere chance. To determine the importance of this correlation, the strength and direction were considered. The correlation was negative, which was not a desired effect. The desired effect of giving procedural feedback is classroom management would improve thereby creating more conducive classroom environments. The correlation was also relatively small. This showed that while there was a significant relationship between procedural feedback and student achievement level outcomes on the FCAT 2.0 Mathematics assessment, it was not a strong relationship.

Summary

In this chapter, data were analyzed to answer five research questions associated with teachers who received specific targeted feedback as defined by Rafalski (2015) and student achievement level outcomes as it related to FCAT 2.0. Observation comments were analyzed to determine categories of specific targeted feedback, and Pearson's r was calculated to examine the relationship of the content of feedback to student success on reading, mathematics, and reading retake assessments. Marzano elements scored during

observations were calculated to examine the relationships to student achievement level outcomes. Data were also analyzed using descriptive statistics and frequencies to determine where potential knowledge voids existed when teachers received feedback.

Research question 1 sought to calculate the frequency of level 1 and level 2 students connected to teachers who received specific targeted feedback. The significance of this research question was to show how many students scored below proficient on FCAT 2.0 in relationship to teachers who predominantly received specific targeted feedback according to Rafalski's (2015) study. Data indicated that of the teachers who received specific targeted feedback, the majority of students scored at achievement level 1 or level 2 on FCAT 2.0 Reading (56.4% of students assessed on reading), FCAT 2.0 Mathematics (59.2% of students assessed on mathematics), and FCAT 2.0 Reading Retake assessments (68% of students assessed on the reading retake assessment).

The purpose of research question 2 was to determine if any relationships existed between the elements scored during teacher observations and mean student achievement level outcomes as measured by FCAT 2.0. This research question was important to determine which elements in the Marzano protocol produced the most student achievement. Frequencies were first calculated and data indicated the majority of elements (14.4%) were coded under element 1 for providing clear learning goals and scales. Many design questions contained minimal comments and some elements (4) contained no comments. Design question 8 only received a total of 0.8% of comments and design question 9 only received a total of 0.6% of comments. These two design questions received the lowest amount of specific targeted feedback.

Pearson's r was calculated to examine relationships between specific elements and student achievement level outcomes as measured by FCAT 2.0 assessments. Calculations revealed only five elements contained any statistically significant relationships with student achievement. Elements 3, 10, 34, 37, and 38 showed significant positive or negative correlations to mean student achievement levels. However, the lack of data in the sample may have affected the results.

Element 3 showed a small, positive correlation in relationship to mean reading achievement levels and a small, positive correlation in relationship to mean mathematics achievement levels. Element 10 showed a small positive correlation to reading retake achievement levels. Element 34 showed a medium, positive correlation to reading retake achievement levels and a medium, negative correlation to mathematics achievement levels. Element 37 showed a small, negative correlation to mathematics achievement levels. Finally, element 38 showed a medium, positive correlation to reading retake achievement levels and a small, negative correlation to mathematics achievement levels. The significance of all correlations existed at either 95% or 99% confidence.

Research question 3 used descriptive statistics to identify the frequencies of certain categories of specific targeted feedback. The feedback was organized into categories based on a rubric. The data showed the predominant type of feedback given consisted of coaching tips (59.0%). Coaching related feedback gave teachers suggestions, recommendations, or ideas to try in their classrooms. Neither the quality nor the validity of the coaching tips was evaluated. The frequency percentages of feedback

categories were as follows: Procedural (20.2%), pedagogy (13.1%), professional development (6.2%), and content (1.6%).

Research question 4 analyzed the charge of feedback to determine if positive, negative, or neutral feedback had an impact on student assessment scores. First, feedback was categorized into specific categories then the charge of the feedback was determined based on a rubric. Positive feedback contained a form of praise or positive recognition, negative feedback focused on what the teacher did not do correctly or well, what the students did incorrectly, or what the teacher should not do. Neutral feedback listed recommendations or suggestions without positive or negative connotations.

Frequencies were calculated and data from this question indicated the predominant charge of feedback given was neutral (50.3%), negative feedback was the next most frequent form of feedback (32.6%), and positive feedback had the lowest frequency (17.1%).

Pearson's r was calculated to examine the significance of relationships between the charges of feedback given to teachers during classroom observations and mean student achievement level outcomes from FCAT 2.0 assessments. According to the data, there were no statistically significant relationships between the charge of feedback and student achievement levels.

While research question 3 identified the frequencies of feedback categories, research question 5 sought to determine if there were relationships between the content of the specific targeted feedback and student achievement level outcomes as measured by FCAT 2.0. Pearson's r was calculated to examine the significance of relationships

between the categories of specific targeted feedback and student achievement level outcomes as measured by FCAT 2.0. The only category of feedback that had any statistically significant relationships was procedural feedback. Procedural feedback showed statistically significant relationships with the mean reading retake achievement level outcomes and mean mathematics achievement level outcomes. There was a small, positive correlation between procedural feedback and reading retake achievement levels. A small, negative correlation existed between procedural feedback and mean mathematics achievement levels.

The findings from the qualitative and quantitative data analyses showed inconsistencies between recommendations in the review of literature as it related to instruction and feedback. Chapter 5 will present and analyze the qualitative and quantitative data from the teacher observation comments and student achievement level data. Chapter 5 will also present a summary of findings; include discussions of the findings, implications for practice, and recommendations for further research.

Table 13

Research Questions, Variables, Sources of Data, Methods of Analysis, and Results

Research Questions	Independent Variable(s)	Dependent Variable	Source(s) of Data	Method(s) of Analysis	Results
1. What is the frequency of level 1 and 2 students in relationship to teachers who received specific targeted feedback?	Teachers who received specific targeted feedback. Historical student assessment (FCAT 2.0) data from the 2013-2014 school year.	Frequency of level 1 and level 2 students	iObservation comments Historical student assessment data	Descriptive statistics: frequency distributions	FCAT 2.0 Reading: Level 1 or 2: 5,507 (56.3%) FCAT 2.0 Mathematics: Level 1 or 2: 5,182 (59.2%) FCAT 2.0 Reading Retake: Level 1 or 2: 236 (68.0%)
2. What relationship, if any, exists between the specific targeted feedback as measured by elements scored during a school year and student assessment outcomes as measured by FCAT 2.0?	Teachers who received specific targeted feedback. Observation data for the 2013-2014 school year. Specific elements scored during observations.	Historical student assessment (FCAT 2.0) data from the 2013-2014 school year. Student assessment outcomes	iObservation comments Historical student assessment data	Descriptive statistics Pearson's r	Element 3 (celebrating success) Mean reading achievement level: $r(89) = .244, p < .05$ (small and positive). Mean mathematics achievement level: $r(89) = .208, p < .05$ (small and positive). Element 10 Processing new information) Mean reading retake achievement level: $r(89) = .229, p < .05$ (small and positive) Element 34

Research Questions	Independent Variable(s)	Dependent Variable	Source(s) of Data	Method(s) of Analysis	Results
					<p>(applying consequences for lack of adherence to rules and procedures)</p> <p>Mean reading retake achievement level: $r(89) = .515, p < .01$ (medium and positive). Mean mathematics achievement levels: $r(89) = -.330, p < .01$ (medium and negative).</p> <p>Element 38 (displaying objectivity and control) Mean reading retake achievement level: $r(89) = .295, p < .01$ (medium and positive). Mean mathematics achievement levels: $r(89) = -.209, p < .05$ (small and negative).</p>

Research Questions	Independent Variable(s)	Dependent Variable	Source(s) of Data	Method(s) of Analysis	Results
3. What is the frequency by content of feedback, defined as content related feedback, pedagogy related feedback, coaching feedback, or professional development feedback provided by observers to teachers during classroom observations?	Observation data for the 2013-2014 school year. Teachers who receive specific targeted feedback.	Types of specific targeted feedback	iObservation comments	Descriptive statistics: frequency distributions	Coaching tips: 266 (59.0%) Procedural: 91 (20.2%) Pedagogy: 59 (13.1%) Professional development: 28 (6.2%) Content: 7 (1.6%) Neutral: 227 (50.3%) Negative: 147 (32.6%) Positive: 77 (17.1%)
4. What relationship, if any, exists between positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0?	Observation data for the 2013-2014 school year. Teachers who received specific targeted feedback. Historical student assessment (FCAT 2.0) from the 2013-2014 school year.	Types of specific targeted feedback Student assessment outcomes	iObservation comments Historical student assessment data	Descriptive statistics: frequency distributions Pearson's r	No significant relationship between feedback charge and student achievement levels.

Research Questions	Independent Variable(s)	Dependent Variable	Source(s) of Data	Method(s) of Analysis	Results
5. What relationship, if any, exists between the content of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0?	Observation data for the 2013-2014 school year. Teachers who received specific targeted feedback.	Student achievement outcomes	iObservation comments Historical student assessment data	Pearson's r	Coaching tip: No significant relationship Content: No significant relationship Procedural: No significant relationship between feedback and mean reading achievement level. Mean reading retake achievement level: small positive correlation $r(89) = .243, p > .05$. Mean mathematics achievement level: small negative correlation $r(89) = -.247, p > .05$. Professional development: No significant relationship Pedagogy: No significant relationship

CHAPTER 5 SUMMARY

Introduction

The preceding chapter contains data and analyses for research questions related to feedback and student achievement. Chapter 5 presents a summary of the study and data analyses, a discussion of the findings, implications for practice on education, recommendations for further research, and conclusions. The summary of this section includes the problem of practice, purpose of the study, guiding research questions, a review of the study's significance and theoretical framework, and the methodology of data collection. The purpose of discussions, implications, and recommendations for further research is to continue to explain the effects of teacher observations on teacher efficacy and student achievement. This section seeks to explain how to potentially improve upon the practice of teacher observations in order to make a positive difference in student achievement.

Summary of the Study

There has been little research on the content of specific feedback given to teachers and how it relates to student achievement. The purpose of this study was to further examine the actual content of feedback to teachers who predominantly received specific targeted feedback as defined by Rafalski (2015) and determine how the feedback related to student achievement level outcomes as defined by FCAT 2.0. The purpose of relating the feedback to student assessment scores was to determine if feedback phrased a specific

way had the potential to increase student achievement as well as determine if the feedback given to teachers helped address knowledge voids in students who scored at achievement levels 1 and 2 on FCAT 2.0.

Bandura's social cognitive theory examines self-efficacy. Self-efficacy is determined by how successful a person believes they will be. The more success a person experiences, the more their self-efficacy increases whereas the more failures a person experiences the more self-efficacy decreases. If a person does not believe he or she will be successful then it is not likely a new behavior will be attempted (Bandura, 1977). Feedback is important to help teachers improve their instructional practices, but if a teacher does not believe in his or her ability to apply feedback to instructional behaviors then the feedback may be avoided. Giving feedback to teachers through the iObservation platform is a means to keep teachers accountable for student achievement with the expectation that the feedback will be utilized to improve classroom practices.

Accountability is a concept that dates back to the 1800s and is written in the current Florida educational code (Florida State Statute § 1012.34). The purpose of accountability is to hold school systems responsible for student progress and success. One way to monitor for this is to observe and record teachers' instruction. Teacher evaluation is a mandatory practice in Florida. Florida statutes require teacher progress to be monitored and 50% to be linked to student data (Florida State Statute § 1012.34(1)(a)). In Large Urban School District, the Marzano protocol is used to determine if teachers are correctly using specific instructional strategies. Teachers are rated as "not using," "beginning," "developing," "applying," or "innovating" depending on the correct usage

of the strategy and successful monitoring of students (Marzano, 2012). As observers are evaluating teacher instruction, they are able to leave feedback in the comments section as it relates to what the observer is seeing. Implications of teacher evaluations may include monetary bonuses in the form of merit pay and job retention or dismissal. Rafalski's (2015) study sought to determine if teachers were receiving specific targeted feedback to help guide their instruction and if the quality of the feedback had a significant effect on teacher VAM (value-added model) scores.

In Florida, during the 2013-2014 school year, the value-added model (VAM) scores were calculated by an algorithm that took into consideration specific variables to produce a score that either met or fell below expectations based on student assessment scores. Results from the Rafalski (2015) study showed that only a small percentage of teachers received predominantly specific targeted feedback and that there was no significant difference in VAM scores that could be attributed to the quality of feedback given to teachers.

In order to narrow the focus on how feedback affects student achievement, the category and charge (neutral, negative, positive) of feedback were identified and analyzed in relationship to student achievement levels on the FCAT 2.0 Reading, Reading Retake, and Mathematics assessments. During the 2013-2014 school year, the results of testing were used in different ways. For example, FCAT 2.0 was the measure of student achievement used in grade 10 to determine if a student graduated high school. It also helped identify students who needed remedial reading or mathematics courses. While different developmental scale score levels were assigned for each grade level,

achievement levels were assigned 1-5 for elementary, middle, and high school students. Therefore, the common achievement level measures were used in student data analyses for this researcher's study. Different categories of specific targeted feedback and feedback charges were also identified for this study. The categories of feedback measured were: Coaching tips, content, procedural professional development, and pedagogy. The charges of feedback measured were neutral, negative, and positive. The criteria for the different categories of feedback and feedback charges can be found in appendices D and E.

Teacher observation data were obtained from a large urban school district and were purposefully chosen based on a sample from a previous study (Rafalski, 2015). The previous study coded feedback from 2,718 teachers to determine if teachers were receiving specific targeted feedback. To determine the effectiveness of specific targeted feedback, a purposive sample of data were obtained from the large urban school district of this study. Teacher's observation comments that were coded as having received specific targeted feedback from Rafalski's (2015) study were matched with their students' assessment scores from FCAT 2.0 and analyzed to determine if statistically significant relationships existed.

The iObservation data for 91 teachers and the assessment scores of their students were analyzed in this study. While it can be noted the teachers in the study were limited to elementary, middle, and high school teachers, this was not specifically delineated in this study. Student outcomes consisted of achievement level data from FCAT 2.0. Assumptions were made based on FCAT 2.0 achievement levels rather than

developmental scale scores (DSS). The reasoning behind this was because scale scores differ amongst grade levels while achievement levels are consistently scored on a 1-5 scale with levels 3, 4, and 5 defined as proficient and levels 1 and 2 defined as below proficient. It should be noted that the bulk to state testing, including FCAT 2.0, was administered after the majority of teacher observations were completed.

This study consisted of both quantitative and qualitative analyses. A qualitative approach was taken to identify what types of specific targeted feedback were being given to teachers as well as the charge of the feedback. Quantitative analysis consisted of converting qualitative data into percentages and calculating Pearson's r to determine correlations between feedback and student achievement.

This study included five research questions:

1. What is the frequency of level 1 and 2 students in relationship to teachers who received specific targeted feedback?
2. What relationships, if any, exist between the specific targeted feedback as measured by elements scored during a school year and student achievement outcomes as measured by FCAT 2.0?

H_{01} . There are no significant relationships between the type of element scored during observations during a school year and student achievement outcomes as measured by FCAT 2.0.

3. What is the frequency by category of feedback, defined as content related feedback, pedagogy related feedback, procedural related feedback, coaching

related feedback, or professional development related feedback provided by observers to teachers during classroom observations?

4. What relationships, if any, exist between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0?

H₀₂. There are no significant relationships between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0.

5. What relationships, if any, exist between the categories of specific targeted feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0?

H₀₃. There are no significant relationships between the categories of specific targeted feedback and student achievement outcomes as measured by FCAT 2.0.

Research question 1 consisted of quantitative data and was answered using frequencies. Research question 2 consisted of quantitative analyses and was answered using frequencies and Pearson's r. Data were calculated into percentages for each teacher and then each teacher's mean student achievement levels were calculated for each test with data. Pearson's r was then calculated and showed if there were any statistically significant relationships between elements coded in observations and student achievement levels as measured by FCAT 2.0. Research question 3 was answered using both qualitative and quantitative analyses. Data were first coded using a rubric, converted to percentages for each teacher, and then calculated to identify frequencies. Research question 4 was answered using frequencies and Pearson's r. Qualitative and

quantitative analyses were used to answer this question. Qualitative analysis included using a rubric to code data based on its charge. Quantitative analysis included converting coded data into percentages for each teacher and then each teacher's mean student achievement levels were calculated for each test with data. Pearson's r was then calculated to determine if statistically significant relationships existed between feedback charge and student achievement levels. Research question 5 was answered using both qualitative and quantitative analyses. Qualitative analysis included coding observation comments with a rubric, and then data was quantitatively analyzed by converting data into percentages for each teacher. Pearson's r was then calculated to determine if statistically significant relationships existed between categories of feedback and student achievement levels as measured by FCAT 2.0.

Discussion of the Findings

Research Question 1 What is the frequency of level 1 and 2 students as instructed by teachers who received specific targeted feedback?

Descriptive statistics and frequencies were calculated from 2013-2014 student FCAT 2.0 data. The student data consisted of 18,875 total assessment scores and achievement levels from 10,745 students. These students were attached to the 91 teachers that Rafalski (2015) found to receive specific targeted feedback from observation comments in a previous study. In the analysis of data, level 1 and level 2 students were grouped together because level 1 and level 2 achievement levels were considered not

proficient, while achievement levels 3, 4, and 5 were considered proficient scores on FCAT 2.0.

Rafalski's (2015) definition of specific targeted feedback stated, "The observer leaves differentiated and meaningful statements intended to improve the impact of an instructional strategy" (p. 79). Data from these results indicated that even though teachers received specific targeted feedback, their students scored predominantly in levels 1 or 2 on the FCAT 2.0 reading assessment, mathematics assessment, and reading retake assessment. These data indicated there were 5,182 students, or 48.2% of the total student sample, who scored at a level 1 or 2 on the FCAT 2.0 mathematics assessment. These data also showed 5,743 students scored at a level 1 or 2 on the FCAT 2.0 Reading and Reading Retake assessments for a total of 53.4% of the total student sample. It is also important to remember that the FCAT 2.0 Mathematics assessment only assessed grades 3-8 and there were no FCAT 2.0 Mathematics Retake assessment scores for this sample of teachers and students. This could explain differences in outcomes when comparing reading to mathematics in this study. While these teachers were coded by Rafalski (2015) as having received specific targeted feedback, they still overwhelmingly had large numbers of level 1 and level 2 students.

For comparison purposes, table 14 showed the percentage of students in Large Urban School District who were level 1 and level 2 students on the FCAT 2.0 Reading assessment. Table 15 showed the percentage of students who were level 1 and level 2 students on the FCAT 2.0 Mathematics assessment. There is no table for the FCAT 2.0 Reading Retakes assessment, only a percentage of students who passed the assessment.

Table 14

District percentages of Level 1 and Level 2 Students on the 2013 Reading FCAT 2.0

Grade Level	Level 1 Students (%)	Level 2 Students (%)
Grade 3	17	25
Grade 4	15	24
Grade 5	15	25
Grade 6	18	22
Grade 7	18	23
Grade 8	16	28
Grade 9	19	29
Grade 10	19	29
Mean	17.13	25.63
Total Level 1 and 2 Students	42.76	

Table 14 showed the percentage of students in Large Urban School District who were level 1 and level 2 students according to the 2013 FCAT 2.0 Reading assessment. According to these data, the average percentage of level 1 students was 17.13% and the average percentage of level 2 students was 25.63%. The total percentage of level 1 and level 2 students was 42.76%. In comparison to the sample used in this study (56.3%), the percentage of level 1 and level 2 students in the district was much lower. The students in the sample of this study had 13.54% more level 1 and level 2 students than the district total.

Table 15

District percentages of Level 1 and Level 2 Students on the 2013 Mathematics FCAT 2.0

Grade Level	Level 1 Students (%)	Level 2 Students (%)
Grade 3	18	24
Grade 4	17	22
Grade 5	21	24
Grade 6	26	24
Grade 7	21	24
Grade 8	22	23
Mean	20.83	23.5
Total Level 1 and 2 Students	44.33	

Table 15 showed the percentage of students in Large Urban School District who were level 1 and level 2 students according to the 2013 FCAT 2.0 Mathematics assessment. According to these data, the average percentage of level 1 students was 20.83% and the average percentage of level 2 students was 23.5%. The total percentage of level 1 and level 2 students was 44.33%. In comparison to the sample used in this study (48.2%), the percentage of level 1 and level 2 students was close, but the amount in the study sample was slightly higher (3.87% difference).

The percentage of students who passed the 2013 FCAT 2.0 Reading Retake assessment in Large Urban School district was 19%. According to state data, 1,937 students took the reading retake assessment. That is a very low percentage rate. In comparison to the results of this study, 347 students took the reading retake assessment and 68% of those students were level 1 and level 2 students. The percentage of students who passed the FCAT Reading Retake assessment in this sample was 32%. While the percentage of students who passed is higher in this sample (13% difference), the scores from this sample consisted of only 18% of all the students who took the retake assessment.

Research Question 2 What relationship, if any, exists between the specific targeted feedback as measured by the type of elements scored during a school year and student achievement outcomes as measured by FCAT 2.0 initial and retake assessments? H_{01} . There is no significant relationship between the type of element scored during observations during a school year and student achievement outcomes as measured by FCAT 2.0 initial and retake assessments.

To answer this question, frequencies were first identified to determine which elements were predominantly scored for teachers who received specific targeted feedback. Data showed there were elements and design questions that received large amounts of feedback and conversely elements that received little to no feedback. The elements that received little to no feedback were of interest because they offer support for engaging low expectancy students. Design question 8 (what will I do to establish and maintain effective relationship with students?) and design question 9 (what will I do to communicate high expectations for all students?) received little to no feedback with their respective elements. The effects of positive relationships may carry into lessons and thus affect student achievement; therefore, more specific targeted feedback could be beneficial in that area.

In relation to design question 8, teacher-student relationships were examined through Hattie's (2009) research. According to a meta-analysis, teacher-student relationships have an effect size of 0.72 (Hattie, 2009, p. 118). "Building relations with students implies agency, efficacy, respect by the teacher for what the child brings to the class (from home, culture, peers), and allowing the experiences of the child to be recognized in the classroom" (Hattie, 2009, p. 118). By taking the time to get to know students and showing care and empathy, teachers can build positive relationships with their students and thereby contribute to helping student achieve. Hattie (2009) says, "In classes with person-centered teachers, there is more engagement, more respect of self and others, there are fewer resistant behaviors, there is greater non-directivity...and there are higher achievement outcomes (p. 119). The data from research question 1 showed that

even with specific targeted feedback that there were many more level 1 and level 2 students than those who scored proficiently. Research question 2 highlighted the gaps associated student performance and instructional strategies as they were observed by school personnel.

Design question 8 is centered on building up level 1 and 2 students by building relationships with students. According to Marzano (2012), those elements are usually included in every lesson and can include teachers not taking behavior issues personally, being able to joke with students, and the ability to have friendly discussions with students.

Design question 9, communicating high expectations for all students was also an area that only received three comments that were coded as specific targeted feedback with no comments being targeted toward element 39-demonstrating value and respect for low expectancy students (Marzano, 2011). Teacher expectations have an effect size of 0.43 (Hattie, 2009, p. 121). The design question focused on how high expectations are communicated to all students. Hattie (2009) references Dusek and Joseph (1985) about the acceptance that teachers "...form expectations about student ability and skills and that expectations affect student achievement" (p. 121). It is how the teachers target the low expectancy students that may make a difference in level 1 and level 2 achievement outcomes.

Strategies such as showing the same positivity for low and high expectancy students, asking questions of the same difficulty and same frequency as high expectancy students, and continuing to probe low expectancy students for correct information are

research-supported best practices (Marzano, 2012). Teachers must hold all students to high expectations in order to help them achieve success. With design question 9 having so little specific targeted feedback, it was speculated that observers may not have seen the interactions, been able to identify the low expectancy students, or scored those elements but did not use specific targeted feedback. It was also speculated based on the number of level 1 and level 2 students, that design questions 8 and 9 may have been overlooked by teachers when planning and teaching lessons, so there were not many opportunities to score those elements. Regardless of the reason, the elements that would specifically target level 1 and level 2 students were not frequently coded and it is clear that if feedback was given, it was not effective enough to help the low expectancy, or level 1 and level 2 students.

Research question 2 also determined if there were correlations between the elements scored during observations and student achievement levels on FCAT 2.0. The results from this test indicated that overall there were very small correlations, most which were not significant. There were a few elements flagged as significant at the 0.01 and 0.05 levels. The following elements showed statistically significant relationships with one or more FCAT 2.0 assessments: Element 3 (celebrating success), element 10 (elaborating on new information), element 34 (applying consequences for lack of adherence to rules and procedures), element 37 (using verbal and nonverbal behaviors that indicate affection for students), and element 38 (displaying objectivity and control).

Marzano (2012) discussed possible strategies teachers could use for each element in his book, *Becoming a Reflective Teacher*. Element 3 (celebrating success), which

showed small, positive correlations in relationship to the FCAT 2.0 Reading assessment and FCAT 2.0 Mathematics assessment, had the following strategies listed: “Final status celebration”, “knowledge gain celebration”, and “verbal feedback” (Marzano, 2012, p. 96). Final status celebration occurs when the teacher celebrates the final outcome of a summative assessment where students made progress. Knowledge gain celebration occurs when students make progress on a learning goal, and verbal feedback consists of teachers recognizing student effort and specifically explaining what the student did well (Marzano, 2012). Examples of phrases for verbal feedback consist of, “You tried very hard on this—good job,” “You were well prepared for this; keep up the good work,” (Marzano, 2012, p. 96) as well as other phrases praising student work and effort. This is a perfect example of how self-efficacy can be increased. If students are seeing positive results and are able to feel like they are making progress then they may continue work hard and make progress. The question then becomes if teachers are asked to celebrate student success as a best pedagogical practice, then should it not be best practice to celebrate teacher success by giving positive feedback?

Element 10 (elaborating on new information) showed a small, positive correlation in relationship to the FCAT 2.0 Reading Retake assessment. Marzano (2012) suggested strategies that require students to think from different perspectives, work in collaborative groups, teach one another, and work with examples of concepts. If examples and collaboration are supposed to help increase student achievement, then teachers would most likely also benefit from collaborating with one another as well as given specific examples of concepts when being observed. It is important to remember that this element

was only scored seven times in relationship to specific targeted feedback accounting for only 1.6% of total comments in this study; therefore, there does not appear to be enough information to accurately correlate this element with student achievement.

Element 34 (applying consequences for lack of adherence to rules and procedures) showed a large, positive correlation in relationship to student achievement on the FCAT 2.0 Reading Retake assessment, but a medium, negative correlation in relationship to student achievement on the FCAT 2.0 Mathematics assessment. It was unclear why the correlation would be positive in relationship to one type of assessment but negative in relationship to another. The researcher predicts this was based on the type of students who generally take the FCAT 2.0 Reading Retake assessment. This element is a procedural element, which requires a specific amount of structure. Students may need to retake the FCAT 2.0 assessment for many reasons such as having not taken the assessment before, missing the assessment, or not passing the assessment previously. Procedural feedback to increase the structure in the classroom could help a teacher increase his or her ability to teach content. There is also the possibility that students who need to take the FCAT 2.0 Reading Retake assessment have the desire to pass in order to graduate if they did not pass in grade 10. It is also important to remember that the cause of a negative correlation could be attributed to the fact that only 16 comments were coded for this element, which is a total of only 3.5% of all comments in this study. Therefore, there does not appear to be enough information to accurately correlate this element with student achievement.

Research-supported strategies for element 34 include ways for teachers to handle students who are misbehaving (Marzano, 2012). The strategies were designed to not disrupt class, teach students consequences, reward students for positive behavior, include the involvement of students and parents, and plan for disciplinary situations (Marzano, 2012). This element had the largest positive correlation with 99% confidence in relationship to student achievement on the FCAT 2.0 Reading Retake assessment. This indicated that student behavior and accountability for their behavior is an important component as element 34 speaks to classroom management. If a teacher cannot control his or her class then it would be difficult to try to teach content; for example, if a teacher is trying to teach a lesson and has to continuously stop the lesson to correct behaviors then students have not learned anything in that class period. If classroom management remains poor all year then that is essentially a year those students have gone without learning the content in that class period.

Element 37 (using verbal and nonverbal behaviors that indicate affection for students) showed a small, negative correlation in relationship to the FCAT 2.0 Mathematics assessment. Some strategies identified as best practice by Marzano (2012) focus on teachers developing positive relationships with students, showing interest, respect, and affection for students, and creating an environment that make students feel valuable. The negative correlations associated with this element are confounding. These types of behaviors should help build students' self-efficacy and make them feel comfortable in their learning environment, which would likely lead to allowing the student to be more open to learning. These behaviors also would appeal to the at-risk

students who may have trouble learning due to low motivation. The negative correlations may be attributed to the possibility that there is only one score coded for this element that accounts for 0.2% of all the comments in this study. Therefore, there does not appear to be enough information to accurately correlate this element with student achievement; however, because of the importance of this nature of element, this researcher suggests that administrators pay attention to the relationships teachers build with students.

Element 38 (displaying objectivity and control) showed a medium, positive correlation in relationship to the FCAT 2.0 Reading Retake assessment and a small, negative correlation to the FCAT 2.0 Mathematics assessment. Some strategies identified as best practice by Marzano (2012) include a focus on the teacher's ability to remain calm in emotional situations, reflect on his or her own behaviors, and actively listen to and effectively communicate with students. This element also speaks to building effective relationships with students, which based on the research of Bandura (1977) and Hattie (2009) this researcher believes all students especially level 1 and level 2 students need in order to be successful in the classroom. It is important to remember that this element was only scored two times in relationship to specific targeted feedback, which accounted for only 0.4% of the total comments in this study. Therefore, there does not appear to be enough information to accurately correlate this element with student achievement. It is of interest that the elements that were found to have significant correlations had such little data. While the amount of data leaves the validity of the correlation in questions, it is noteworthy to point out that the element with the most comments, element 1 (providing clear learning goals and scales) showed, no significant

correlations. This information could help observers prioritize which elements are coded during observations and lead to increasing the frequency of feedback given in relationship to these elements.

This study showed there were small correlations with all elements and student achievement, which was predictable based on the nature of social science. It is of interest that there were only five elements that showed any sort of significant relationship to student achievement. It is unclear why those elements showed significant correlations with so little data in terms of how often the elements were scored. What is important to note about the significance of those correlations is the importance of those elements and design questions in regard to the impact they have specifically on level 1 and level 2 students. This researcher recommends that administrators weigh the significance of teachers building relationships with students, showing high expectations for all students, and requiring students to think at higher levels of rigor when conducting classroom observations.

Research Question 3 What is the frequency by content of feedback, defined as content related feedback, pedagogy related feedback, procedural feedback, coaching feedback, or professional development feedback provided by observers to teachers during classroom observations?

This research question was answered using qualitative coding and frequencies. Data were first coded using a rubric (located in appendix d) and then rechecked to determine reliability. Once the data were all coded, frequency distributions were calculated to determine how often each of the content codes occurred during classroom

observations where specific targeted feedback was identified. It is important to note that many comments could have been placed in multiple categories; however, comments were only placed into the category that showed predominant traits according to the rubric in appendix d.

Based on the frequency distribution, 266 (59.0%) comments, which was the majority of comments, were coaching tips. These coaching tips consisted of administrators giving teachers feedback on what could be done differently in the class or what else could have been done in the class in order to have the most success within that element. Coaching tips included examples and recommendations for teachers to try. What was missing in the comments was follow-up on how teachers may have performed on trying the coaching tip. There were very minimal circumstances where an administrator gave a coaching tip for teachers to try and then followed up to see if the teacher applied the feedback and what successes resulted from trying the new strategy. In the circumstance a coaching tip is monitored by an observer, the same elements should be tracked in order to ensure growth with the specific teaching strategy. An interesting observation about the coaching tips was there were no comments that indicated the teachers should seek out the aid of the school based coach nor were there any comments that indicated a school based coach had been working with a teacher in the past or would be working with a teacher in future.

The next most frequent type of feedback was procedural with 91 (20.2%) comments. This type of feedback included administrators making comments or giving tips specifically related to classroom rules and procedures. Examples of this feedback

included comments related to classroom layout, restroom procedures, management procedures, following the rules, and student behavior. Procedural feedback is important because it speaks to classroom management. Classroom management must occur in order for teaching and learning to occur. If a teacher spends most of his or her time trying to enforce rules and procedures then instructional time is lost and no number of instructional strategies can make up for the lost time. Setting the tone and structure for what is acceptable behavior and what specific rules and procedures are for the class should occur at the beginning of the year and be practiced until the behaviors become routine. As the instructional strategies on the Marzano learning map show, there should also be consequences for students who do not follow class rules and procedures. The feedback related to procedures in this study did in fact include comments on what teachers should do to improve student behavior.

Pedagogical feedback consisted of 59 (13.1%) comments. The feedback for this category of feedback consisted of observers commenting on Marzano strategies and how they were used as well as how the teacher monitored students. Pedagogy feedback also consisted of observers telling teachers how they could move from one level to another, for example from developing to innovating. The researcher coded this feedback as pedagogy since the comments about the teaching practices aligned with the evaluation framework used in the large urban school district of this study. The quality of the comments were beyond the scope of this study, but as an observation from reading the comments, this researcher noticed some feedback that was helpful and some feedback that simply told a teacher to continue working on an instructional strategy. In order for

feedback to be helpful, it should include what the observer saw, what worked, how the instructional strategy could be improved upon, and how to effectively monitor for the desired effect of the element coded. Many of the feedback comments told teachers to change or correct a strategy but did not coach the teacher on how it could be done.

Professional development feedback consisted of 28 (6.2%) comments. This category of feedback consisted of comments that referred teachers to professional development resources such as meeting with another teacher or referencing a type of literature such as the Marzano text, *Becoming a Reflective Teacher* (2012) or *The Art and Science of Teaching* (Marzano, 2007) . The quality of feedback related to professional development was beyond the scope of this study; however, from reading the comments there was no evidence of follow-up to professional development comments. It was unclear if the administrator had previously suggested professional development and there were no comments that indicated any improvement was observed as a result of professional development. Professional development should be specific to teacher needs and include collaboration. Referencing a teacher to a resource to read about a strategy he or she is struggling with may provide more information about the strategy but does not give any practical credence to increasing instructional practices. Overall, the information in the literature would theoretically improve the teacher's knowledge of specific instructional strategies provided the teacher actually referenced the materials. For professional development to be effective though it should be relevant to the teachers' individual needs and monitored for effectiveness.

Content related feedback received the fewest amount of feedback comments at 7 (1.6%) out of 451 comments. This feedback consisted of comments that specifically related to the content being taught. While many comments referred to the subject area, only the 7 comments coded consisted of feedback that supported or enhanced the content being taught. This support included specific examples of how a strategy could work in a particular content area while using the information observed being taught in the class. Observers in Large Urban School District may observe teachers in a subject area for which they are not content experts. For example, an observer may be required to observe a science teacher but never taught science and therefore not be proficient in science standards or content. As such is the case, it may make observers uncomfortable to comment directly on how a teaching strategy supports or can enhance a content area standard. It is the opinion of this researcher that teachers who are consistently teaching a subject area may feel they are more qualified to assess the content than an administrator who has not taught that particular subject area or has not taught in many years.

In the large urban school district of this study, observers currently using the Marzano protocol must now conduct observations to include content area standards knowledge as part of the observation process. The current protocol only requires observers to determine if teachers are using instructional strategies correctly and monitoring for their desired effects. The new protocol will not only require administrators to determine if instructional strategies are being used correctly and monitored for the desired effects, but they must be able to relate these strategies to the content area standards. If a teacher is using a strategy correctly but did not teach toward

the full intent of the appropriate standard then the teacher could not possibly score higher than “beginning” according to the revised protocol. The rationale behind this change was if a teacher was identifying critical content (element 6) for example, but the content they were teaching was not standards-based, then the teacher was essentially teaching wrong information no matter how beautiful the presentation of the lesson appeared. It is important for observers to be able to give teachers feedback on the content they are teaching and especially vital to catch mistakes if teachers were not teaching information correctly. Therefore, observers should be able to give content specific feedback that teachers can immediately apply to their own classes. To remedy the issue of observers not having the content area background for the subject they assess, using resources such as content standards, and personnel such other administrators with content knowledge or school coaches can help fill in the knowledge gaps of the administrators in order to deliver more relevant and valid specific targeted feedback.

The data analyzed from this research question has led to many conclusions about the content of specific targeted feedback. First, coaching tips are helpful when teachers are able to put them into practice immediately. Observers must be willing to follow up on the coaching tips given in previous observations on the same elements. Procedural feedback is important because of its implications on classroom management, which teachers need in order to deliver smooth, uninterrupted instruction. Feedback related to pedagogy should not just tell a teacher to continue to use a strategy but tell a teacher what was done well and how the strategy could be improved in addition to monitoring for the desired effect of the element coded. Professional development feedback is not properly

developing teachers because of the loose connection to actual development. Professional development should consist of more than referring teachers to literature in order to read about instructional strategies. Finally, there was a lack of content related feedback, which could suggest a lack of content knowledge of observing administrators. An example of each category of specific targeted feedback is located in appendix f of this study.

Research Question 4 What relationship, if any, exists between the frequencies of positive, negative, or neutral feedback and student achievement outcomes as measured by FCAT 2.0?

H₀₂. There is no relationship between the content of specific targeted feedback and student achievement outcomes as measured by FCAT 2.0.

This research question was answered using qualitative and quantitative measures. Data were first coded using a rubric (located in appendix e) and rechecked to determine reliability. The rubric was used to determine if feedback was neutral, negative, or positive. Data were then calculated using descriptive statistics in order to identify the frequencies of each charge of feedback.

There were comments in this study that had components of multiple charges. An example would be an observer saying something positive such as how the teacher was improving using a strategy but then commenting on what did not work or how students were off task. In these cases, the predominant charge of the feedback was determined based on the ratio of charges within the comments and overall tone of the comment. In the majority of comments, the charge was easily identifiable.

Data from this research question showed the most frequent type of feedback had a neutral charge. Neutral feedback consisted of 227 (50.3%) comments. Generally, this type of feedback consisted of observers telling teachers what was seen during the class, suggestions or recommendations for learning strategies, and what to look for or monitor during lessons. Neutral feedback is safe in the sense it does not incite emotions over observations; however, based on the research of Bandura (1977) and Hattie (2009) some praise is needed in order for teachers to feel validated and increase self-efficacy. While Hattie (2009) does not recommend giving praise alone, feedback should not pose a high threat to self-esteem. By increasing self-efficacy, teachers may be more apt to try new strategies. If a teacher tried a strategy and was not given appropriate feedback which consisted of support and encouraged reflection, then that teacher may not be amenable to future feedback or trying strategies identified as best practice.

Negative feedback had the next highest frequency with 147 (32.6%) comments. This feedback consisted of comments that said students or the teacher were not doing something. For example, the teacher was not monitoring or the students were not on task or did not listen. Not all negative comments were empty criticism. There were comments that gave coaching tips for suggestions to improve. With the implications of teacher evaluations on pay and the suggestion that a teacher is not doing a good job, this researcher wonders how often teachers actually read their feedback and how they react when they are given negative feedback. It was unknown if teachers were being given negative comments because they were consistently underperforming. The question then becomes what kind of feedback should teachers receive who do not follow feedback and

are not using instructional practices as they should be? Consistent observations that highlight deficiencies in instruction are a way administrators may be able to document teachers who are in need of improvement. These teachers should continue to be developed in order to improve their instruction, but how should administrators phrase feedback if nothing positive is observed? In the circumstances where an observer may want to give negative feedback, perhaps keeping feedback neutral is beneficial as there are likely instances where an administrator may not want to give praise to a teacher but still want to help the teacher for the sake of the students.

The lowest frequency of feedback charge was positive feedback with 77 (17.1) comments. Positive feedback consisted of comments that complimented the teacher on the use of a strategy, how the teacher engaged students, or if the teacher had improved. The low amount of positive comments indicated that teachers were not receiving much praise in what they were doing correctly in their instruction. Even much of the praise teachers received did not come without the observer commenting on how a strategy may not have been as effective as it could have been. An example of this would be an observer telling the teacher the strategy used was a great strategy but students were still not monitored enough or were unclear of information. The research on feedback by Hattie (2009) does say that there should be a multilayered approach to giving feedback that requires teachers to be reflective.

After the frequencies were analyzed the qualitative codes were converted to percentages and Pearson's r was used in order to determine if there were any statistically significant relationships between the charge of feedback given during teacher

observations and student achievement of their students as measured by FCAT 2.0. Data from the Pearson's r calculations indicated there were no statistically significant relationships between the charge of feedback and student achievement outcomes. While there will usually be correlations in the social sciences, the charge of feedback did not show any significant correlations to student achievement levels as measured by FCAT 2.0. Since there were no significant correlations between feedback charge and mean student achievement levels, the null hypothesis must be retained.

Data from research question four had many interesting findings. Frequency data showed the most frequent charge of feedback was neutral and the lowest charge of feedback was positive. Correlational data indicated there were small correlations as expected in the social sciences, but there were no significant correlations meaning the null hypothesis had to be retained. It is possible there were no significant correlations because of the sample size of comments in comparison to the sample size of student achievement outcomes. Examples of comments with each charge of specific targeted feedback are located in appendix f of this study.

Research Question 5 What relationship, if any, exists between the content of targeted specific feedback provided to teachers and student achievement outcomes as measured by FCAT 2.0 initial and retake assessments?

H_{03} . There is no relationship between the content of specific targeted feedback and student achievement outcomes as measured by FCAT 2.0 initial and retake assessments.

This research question was answered using qualitative and quantitative measures. Data were first coded using a rubric (located in appendix d) and rechecked to determine

reliability. The qualitative codes were converted to percentages in order to allow for a correlational analysis in relationship to student achievement levels. The purpose of this research question was to determine if there were any significant correlations between the type of specific targeted feedback (coaching tips, content, procedural, professional development, and pedagogy) and student achievement levels as measured by FCAT 2.0.

Data showed there was only one type of feedback that had any significant correlations with mean student achievement levels. Procedural feedback showed significant correlations in relationship to the FCAT 2.0 Reading Retake assessment and the FCAT 2.0 Mathematics assessment. Procedural feedback refers to rules, procedures, classroom management, or classroom layout. Based on the Steinberg (2011) measurements, the correlations were small.

The correlation between the reading retake assessment and student achievement levels was a small, positive correlation, which indicated a prediction that as procedural feedback increases student achievement on the FCAT 2.0 Reading Retake assessment will increase. As a generalization, students who retake the reading assessment are English Language Learners (ELL), exceptional needs students (ESE), or are at-risk of not graduating. These students more than likely need the structure associated with clear and enforced rules and procedures. As students who may be in remedial reading classes, these procedures are important in building classroom management routines that will allow for teachers to deliver the critical content necessary to help students pass the reading retake assessment and not waste instructional time correcting behavior issues.

The correlation between the mathematics assessment and student achievement levels was a small, negative correlation. This indicated a prediction that as procedural feedback increases student achievement on the FCAT 2.0 Mathematics assessment will decrease. While rules and procedures speak to classroom management and allow for content to be taught, it is possible that different types of students need differentiated focus on rules and procedures. The assumptions about students who take the reading retake assessment may not apply to students who are taking the mathematics assessment for the first time. It is also possible mathematics teachers did not receive as much feedback related to procedures as teachers who taught reading.

This study did not delineate the subject area of the teacher receiving feedback. A possible explanation for why specific assessments were flagged as having significant correlations and coded positive or negative could be a result from the teacher who received the feedback and the observer who gave it. This could explain why feedback affected the reading assessment and not the mathematics assessment. If a remedial reading teacher received specific feedback then that may have affected the scores on the FCAT 2.0 Reading Retake assessment. If an English teacher received feedback, that may have affected achievement levels on the FCAT 2.0 Reading assessment, but not the FCAT 2.0 Mathematics assessment and vice versa.

Since this study analyzed all student data associated with teachers who received specific targeted feedback, all students connected to those teachers who took FCAT 2.0 were included. There was also no indication as to which schools were included in this study. The schools in the large urban school district of this study included both low and

high performing schools with varying levels of students and parental involvement. As a generalization, the lower performing schools may have had more issues with behavior and therefore teachers may have needed more feedback on procedures and classroom management.

This research question showed there were no significant correlations for all but one category of specific targeted feedback: procedural. It was unexpected that there were no other significant correlations to the types of feedback given, which raised the question about how often teachers were reading and applying the feedback given to them during observations.

Procedural feedback is important because of implications in classroom management. The structures needed for specific students may account for the positive significance of feedback to student achievement levels on the FCAT 2.0 Reading Retake assessment. Regardless of the class, all teachers should have procedures in place in order to be able to teach the content in their classrooms without constant disruptions. Once the procedures are in place, teachers can then focus on how to refine instructional practices and strategies assuming they are reading and applying the feedback from their observations. Procedural feedback was the only category that showed any significant correlations to student achievement. This suggests that the ways in which feedback is being delivered is not as effective as it should be in increasing student achievement.

Implications for Practice

The purpose of teacher observations as they were used in the large urban school district of this study was to help teachers improve upon their teaching practices in order to increase student achievement. The Marzano instructional framework and VAM were used in the school district of this study as a means of evaluating teachers' impacts on student performance.

The findings of this study had many implications for the school district and school administrators in order to deliver classroom observations that would be more effective for teachers and student achievement. Findings from this study should give school administrators and district personnel ideas on how to frame feedback for teachers, what specific attention to pay to feedback published to teachers and to the school district, what to focus on during classroom observations, how to follow up and support teacher accountability for specific improvement, and what areas could help improve student achievement specifically for level 1 and level 2 students.

Accountability is a general term used to hold schools, teachers, administrators, and the school district responsible for student achievement. This study focused on accountability as measured through a combination of teacher observations and student achievement outcomes; it analyzed achievement results from the 2013-2014 FCAT 2.0, a summative reading and mathematics assessment. Feedback given to teachers was also analyzed to determine the category and charge of feedback teachers were receiving during classroom observations. An effective or highly effective evaluation could mean a teacher received more money in the form of merit pay, while an evaluation that indicated

needing improvement could mean a teacher was put on a plan to monitor and improve his or her instructional practices because of potential problems.

This study resulted in six implications for district administrators and school administrators and others who give teachers feedback using the iObservation platform.

- 1) Observation feedback is not significantly affecting student achievement.
- 2) Feedback needs to include a focus on targeting level 1 and level 2 students.
- 3) There should be follow-up to feedback in order to monitor accountability throughout the year.
- 4) Feedback should be clear and immediately applicable.
- 5) There is not a significant amount of feedback targeting higher-level thinking.
- 6) There is a need for more effective communication between teachers and observers regarding implementing feedback.

The first implication that observation feedback was not significantly affecting student achievement was based on the data from research questions 1, 2, 4, and 5. Data from research question 1 showed there were higher frequencies of level 1 and level 2 students than proficient students of teachers who received specific targeted feedback. If specific targeted feedback was significantly affecting student achievement, then theoretically there would be more students scoring proficiently on the summative assessment.

Data from research question 2 showed the relationships between Marzano elements scored and student achievement outcomes. The relationships were sporadic and inconsistent as only five elements showed any statistically significant relationships and

there were different results depending on the tests. This could be attributed to the data set; however, as the Marzano elements are research-supported best practices, there should be more significant relationships between their use and student achievement. Another reason for the inconsistent relationships between the elements and student achievement may be observers' allowing the Marzano protocol to drive observations as opposed to watching instruction and then determining what elements were being used within the lesson. While observers were trained to identify predominant elements while watching lessons, qualitative analysis of the feedback showed comments that were incorrectly coded.

Data from research question 4 showed there were no statistically significant relationships between the charge of feedback and student achievement outcomes. According to the results, it did not matter if feedback was phrased neutrally, negatively, or positively. This countered the research about self-efficacy and praise, which stated self-efficacy increases as the belief of success increases and decreases as failures increase (Bandura, 1977; Owens & Valesky, 2011). While Hattie (2009) said praise alone is not helpful, he gave a framework for how feedback could be delivered in order to ensure reflection is occurring and discussed the importance of feedback threatening self-esteem. The framework incorporated multiple levels such as directions, understanding, and reflection, which could help guide teachers toward what needs to be accomplished. If negative feedback was a threat to a teacher's self-esteem, then the expected results should have shown statistically significant relationships between feedback charge and student achievement: a positive correlation for positively charged feedback and a negative

correlation between negatively charged feedback. In other words, as positive feedback increased then student achievement would increase while as negative feedback increased student achievement would decrease. It is recommended that observers consider how people may react to feedback delivery. If the intent of the feedback is for teachers to read and apply it, it should be delivered in a manner that invites teachers to embrace feedback, not ignore it.

Research question 5 sought to determine if there were relationships between the type of specific targeted feedback provided to teachers and student achievement outcomes. Data showed only one type of feedback (procedural) had any significant correlations, which were small, and it too yielded mixed results. Again, while this may be attributed to the data set, the other feedback types showed no significant correlations suggesting the problem was not only with the data set but also with the effectiveness of the feedback. These data implied that while comments fell into the category of specific targeted feedback as coded by Rafalski (2015) they were not effective in increasing student achievement outcomes.

The second implication: Feedback needed to include a focus on targeting level 1 and level 2 students, was based on data from research questions 1 and 2. Data from research question 1 showed higher frequencies of level 1 and level 2 students than students who performed proficiently on FCAT 2.0. Since the 91 teachers in this study received specific targeted feedback, it was predicted there would be higher student achievement outcomes. These data implied the feedback given to these teachers was not effective in supporting the needs of level 1 and level 2 learners. The data from research

question 2 examined if there were significant relationships between the Marzano elements scored during observations and student achievement as measured by FCAT 2.0 Reading, Reading Retake, and Mathematics assessments. According to the data presented in research question 2, there were statistically significant relationships with only five of the Marzano elements. These elements belonged to a broader category of design questions. Specific design questions were intended to support all students especially level 1 and level 2 learners. These design questions were design question 8 (What will I do to establish and maintain effective relationships with students?) and design question 9 (What will I do to communicate high expectations for all students?). These two design questions received little to no specific targeted feedback comments. The elements within these design questions included specific instructional strategies for engaging low expectancy students, or students who were likely level 1 or level 2 students. Specifically, design question 9 focused on increasing rigor with low expectancy students. In order to increase the achievement of all students, especially level 1 and level 2 learners, it is recommended more feedback be given in these areas to target level 1 and level 2 students.

The third implication suggested there should be follow-up to feedback in order to monitor teacher accountability throughout the year. This implication was based on the qualitative analyses in this study. In order to categorize the iObservation feedback, the comments were read multiple times and scored using a rubric to measure reliability. Each of the 451 comments were examined to determine the category of feedback and the charge of feedback. Reading these comments highlighted that there were little to no

statements about expectations from previous observations or follow-up to previous expectations. This suggested that even if coaching tips or suggestions were given previously, there was no continuity or system in place in terms of ensuring if the feedback was understood and implemented. While suggestions were not mandated, the lack of follow up with teachers may send the message that the feedback comments did not matter. Data from this study implied that follow-up feedback was needed to monitor teacher use and success with specific instructional strategies. When teachers were observed, they were scored on how they monitored student progress toward mastery of standards; teachers also needed to be supported on progress toward correctly using best pedagogical practices in order to increase student achievement.

The clear and immediate application of feedback was the fourth implication in this study. Data to support this implication was highlighted in research questions 3 and 5 as well as in the qualitative data collection and analyses of this study. Research question 3 showed the frequencies of the different types of feedback categories. The highest frequency of feedback was coaching tips (59.0%). Most of the feedback delivered to teachers during observations were suggestions and recommendations about instructional practices. However, data from research question 5 showed there was no significant relationship between coaching tips and student achievement as measured by FCAT 2.0. These data suggest, as mentioned in the third implication, that teachers may not have acknowledged feedback or clearly understood what to do with the feedback.

The qualitative analysis gave more information about the fourth implication in terms of the ability of teachers to clearly understand how to use suggestions and coaching

tips. Each comment was read multiple times using a rubric to measure reliability. Coaching tips, which included recommendations and suggestions, were delivered to teachers via the iObservation platform. There was no written indication that anyone modeled suggested strategies for teachers or met face-to-face to discuss the informal observations. Many of the coaching tips simply gave the tip, but did not tell the teacher how to use the tip or plan with it. For example, an administrator may tell a teacher to consider allowing the students to be more involved in identifying the learning goal during a lesson, but not explain how to use that tip or what that would look like during the lesson. Therefore, it was important for administrators to ensure coaching tips were understandable to teachers to be able to try with success as soon as they were read. Another observation made while reading the feedback comments was the frequency with which there were spelling, grammatical, and syntactical errors in the feedback. It is recommended that observers take care to reread feedback before submitting it to teachers. Appendix H shows examples of feedback comments with spelling, grammatical, and syntactical errors.

Data from research question 2 revealed there was a lack of specific targeted feedback that addressed higher-level thinking. The frequency of Marzano elements for design question 4 (What will I do to help students generate and test hypotheses about new knowledge?) showed only 1 comment (0.2%). These data showed a larger focus on element 1 (Providing clear learning goals and scales) with 65 comments (14.4%) dedicated to this element. Goals and scales may have been useful in helping students to

track the level of rigor necessary for higher-level thinking tasks; however, teachers should have been able to show student outcome improvement from using these elements.

The last implication identified in this study was the need to improve ongoing communication between observers and teachers as it related to feedback. This study showed the categories of feedback and charges of feedback were mostly inconsequential to student achievement outcomes. Procedural feedback was the only category of feedback to show a significant relationship to student achievement outcomes. Data from this study revealed administrators did indeed give feedback to teachers; however, its usefulness remains questionable. While the quality and validity of the feedback were beyond the scope of this study, the dearth of statistically significant relationships between observer feedback and student achievement outcomes suggested feedback was not meeting its intended effect of increasing student achievement outcomes. The qualitative data analysis from this study showed observers made suggestions to teachers to improve upon an instructional practice; however, there was little evidence of observers following up with teachers regarding implementing feedback in their classrooms. This evidence suggested the process of how feedback was delivered and monitored using the iObservation platform may need to be reevaluated.

Accountability is not simply a summative concept that is evaluated once a year, it is a continuous process in which teachers' instruction is continuously monitored. Often teachers may have felt like they only needed to put on a show of compliance for administrators a few times a year then go back to doing what they wanted to do because of the cumbersome observation process. Once the observation period ended, teachers

may have no longer worried about observation compliance. These behaviors should be changed so teachers understand that observations are about the students and not just about compliance. Therefore, it should have been the expectation that if an observer noticed a teacher needed to make an instructional change that the proper feedback and supports were delivered so that teachers could use the information to increase student achievement. Once teachers were given the help they needed, they should be expected to make the necessary changes. This means as observers scored elements; those elements should be revisited in subsequent observations to monitor progress. It is important to remember social cognitive theory (Bandura, 1977) when giving feedback to a teacher trying a new strategy. If a teacher did not use a strategy correctly, an observer should be mindful that more help might be needed in order to master that strategy. It would not be helpful to give negative feedback while a teacher is developing a practice because a teacher may stop trying new strategies. What was needed was specific targeted feedback with clear steps and expectations for practice. There should also be expectations for teachers to take the steps necessary to fill instructional gaps. It was the responsibility of the observer to give feedback that could be clearly understood, implemented, and then monitored to track progress. Observations should not be viewed as punishment, but opportunities for coaching and professional development for the success of all students.

This information should help administrators continue to give specific targeted feedback that is accurate, relevant, timely, and aids in changing teacher behaviors in order to improve student achievement especially for level 1 and level 2 students.

Accountability is not a negative concept: It is in fact important in making sure teachers

and administrators are using best practices to help students succeed. There may be negative connotations associated with the word accountability because of personnel and monetary consequences; however, evaluations are necessary in order to ensure schools are systematically meeting the needs of all students. Observers may be viewed as individuals who simply walked around trying to “ding” teachers on what they were not doing, but it was the job of the observer to collect qualitative data in order to determine the needs of teachers. According to the Florida Department of Education (2012), as of 2012, 44% of Florida school districts were using the Marzano model to observe teachers. The issue was not with the evaluation system existing, but with the process in which the protocol was used. It was speculated the personnel in the large urban school district of this study undoubtedly wanted all students to achieve and succeed. By carefully considering how feedback is delivered to teachers about their instructional practices, fine-tuning the process of writing comments, and focusing on specific areas that could eliminate knowledge voids, it may be possible to increase student achievement.

Recommendations for Further Research

The purpose of this study was to investigate the relationships between specific targeted feedback comments given to teachers during classroom observations and student achievement outcomes as measured by a summative standardized assessment (FCAT 2.0). Data were collected to test five research questions related to this purpose. The information was examined and resulted in some substantial findings from the analyses of

the data. However, there were some limitations to the findings, which influenced the following recommendations for further research.

Data analyses from this study resulted in 11 recommendations for further research:

- 1) Further research which closely examines the contents of feedback for accuracy, clarity, and relevancy could be helpful in determining reasons why there were little to no relationships between observation feedback and student achievement.
- 2) Replication of this study with a focus on one grade level or one education level.
- 3) Replication of this study with a focus on one assessment.
- 4) Future research analyzing how observers are trained to help identify gaps in the observation training processes.
- 5) It may be beneficial for future research to analyze a sample with a larger percentage of specific targeted feedback that focuses on design questions and elements related to helping level 1 and level 2 students.
- 6) Further research could include collecting perception surveys from teachers. Survey data collection could include how teachers feel about observations, feedback, and reasons why feedback is accepted or rejected. Data should also include what would prompt teachers to respond to feedback and change behaviors based on observation feedback as well as data collection on

trainings and professional development teachers need to effectively implement feedback.

- 7) A study on the relationship between school culture and student achievement as the result of feedback.
- 8) Replication of this study in another district using the Marzano protocol could determine if results from this study are a result of how observers are trained in the large urban school district of this study.
- 9) Replication of this study using the new FSA assessment.
- 10) A study on the effects of conducting all observations (informal and formal) using the formative supervision approach and the relationship to feedback acceptance and student achievement.
- 11) Analyze the feedback given to school and district leadership.

One limitation was that there was no consistency between the significance of the results from test to test; this may be in part due to the data set. The data set only included 451 comments related to different Marzano elements, and the comments were not equally dispersed between the elements. For example, one element may have had 12 comments while another element may have had only 2 comments. When relating these data to student achievement scores, there was often not enough information to determine if some relationships were truly significant.

Another limitation was that the qualitative comments were sometimes coded incorrectly by the observers. To answer research questions 3, 4, and 5, the comments in this study were organized using rubrics to determine the category of feedback and the

charge of feedback; however, the validity and relevancy of the feedback were not statistically analyzed. While the comments were read multiple times to determine rubric placement, the research questions did not include further analysis of the extent with which the feedback could be applicable to instructional practices. Further research which closely examines the content of feedback for accuracy, clarity, and relevancy with an equal data set might be helpful in determining reasons why there were often little to no relationships between observation feedback and student achievement.

Implications from the data analyses included the lack of clarity in observation comments and the lack of follow-up on elements scored in successive observations. As observers were commenting on specific elements, other elements may have been discussed or referenced. This is a problem because the teacher was actually getting feedback on an element other than one being coded in the iObservation platform. This could cause confusion to a teacher, give data that was not indicative of what was actually occurring in the classroom, and even influence a teacher to change his or her instructional practices incorrectly. Another implication included the lack of follow-up from one observation to another. It was assumed since the comments were from the 2013-2014 school year that they were from multiple observations. There was a lack of clear information in the comments about monitoring the teachers' progress using specific instructional practices from one observation to the next. The frequency with which observers followed up on feedback given to teachers was not analyzed in this study. Future research analyzing these implications could help identify gaps in the observation training processes for observers.

There were specific assumptions made with the data set in this study. The researcher chose to use proficiency levels as defined by the state as the measure of student achievement. Since elementary, middle, and high school student data were reported differently, achievement levels were used as opposed to developmental scale scores (DSS). DSS scores differed amongst grade levels while achievement levels measured at 1-5 for all grade levels. Students who scored at achievement level 3 or higher in each assessed grade level were considered proficient. For this reason, it was necessary to analyze achievement levels by averaging them for each teacher and comparing them to the average comment category or charge. It is recommended future research consists of data analysis structured by a single grade level (i.e. 10th grade) or education level (i.e. high school).

Another limitation with the data among the grade levels was the FCAT 2.0 Mathematics exam only assessed grades 3-8, the reading exam assessed grades 3-10, and the reading retake exam assessed specific students in grades 11 and 12. In order to get more accurate results with assessments and feedback, the study could be replicated while only focusing on one assessment: reading or mathematics. This may also eliminate some of the inconsistencies with the results between assessments. For example, procedural feedback showed a negative correlation with the FCAT 2.0 Reading Retake assessment but a positive correlation with the FCAT 2.0 Mathematics assessment. Generalizations may be drawn about the students or the data, but perhaps focusing on only one type of assessment or even narrowing the assessment limitations to one level (elementary, middle, or high school) would give more clear and accurate results.

Quantitative data analyses showed a lack of specific targeted feedback related to design questions 8 and 9 that specifically targeted level 1 and level 2 students. Design question 8 and design question 9 included elements on building relationships with students and holding low expectancy students to high expectations. The total percentage of comments made related to these design questions was a mere 1.4%. It may be beneficial for future research to analyze a sample with a larger percentage of specific targeted feedback which focuses on design questions and elements related to helping level 1 and level 2 students. Because of the low amount of comments in this study related to design questions 8 and 9, the results could not be generalized even though two of the elements showed significant correlations. These correlations may not be as significant as suggested by the results from Pearson's r because of the low number of comments calculated with student achievement outcomes (3 total).

Based on the qualitative data analyses from this study, an implication on education was that more effective communication was needed between observers and teachers in terms of implementing feedback. Further research could include collecting perception surveys from teachers and a study on the relationship between school culture and student achievement as the result of feedback.

Conducting research on the school level would potentially allow for more detailed data to support qualitative analyses. Survey data collection could include how teachers feel about observations, the culture and climate of the school in regard to observations, feedback, and reasons why feedback is accepted or rejected. Data could include what might prompt teachers to respond to feedback and change behaviors based on observation

feedback. Further data collection could also include information about trainings and professional development teachers need to effectively implement feedback.

Another limitation of this study was that it was conducted in only one large urban school district. In order to see if the data from this study were valid, a replication of this study from the 2013-2014 school year in another district using the Marzano protocol is suggested. This might provide additional data to better determine if the results may have been influenced by how observers were trained in the district in this study.

Another limitation is that there is no way to examine if the results have changed over the last 2 years. The standardized test, which was used to assess students in Florida has changed. The 2013-2014 school year was the last year in which FCAT 2.0 was used to determine student proficiency in mathematics and reading. Beginning the 2014-2015 school year, the Florida Standards Assessment (FSA) was adopted in addition to new standards aligned with Common Core. In addition to examining another district, it is also suggested that the study be replicated using the new FSA.

In the district used in this study, administrators were also evaluated using the Marzano protocol but with very different elements. Since many administrators are frequent observers of teachers, it would be of interest to examine how they were receiving feedback to help determine if there are any systematic issues with using a single prescribed observation tool, and how it is being used. One last recommendation for future research would be to analyze the feedback given to school and district leadership.

The literature review thoroughly discussed the concept of formative supervision. According to researchers, the purpose of formative supervision is to help build teachers'

knowledge and practices through constant observations and feedback (Nolan & Hoover, 2011; Range, Young, & Hvidston, 2013; Kelting, Jenkins, & Gaudreault, 2014; Tang & Chow, 2006). The current protocol in the large urban school district is to provide teachers with feedback in iObservation for all observations that count toward evaluation and to meet with teachers before and after formal observations. The clinical supervision model requires pre conferences and post-conferences with teachers for informal observations in addition to formal observations. Since data from this study revealed a larger number of students who did not pass FCAT 2.0 assessments than students who scored proficiently, a study on the effects of conducting all observations (informal and formal) using the clinical supervision approach and the relationship to feedback acceptance and student achievement could provide the school district with more information on the potential benefits of using the clinical supervision model for all observations.

These recommendations for future research were based on the data collected and analyzed from this study. The implications of this study also contributed to ideas for future researchers to examine. The purpose of these recommendations was to help future researchers narrow the focus of specific targeted feedback to help level 1 and level 2 students, while also being mindful of how feedback is delivered to teachers.

Conclusions

The findings in this study expanded on the work of previous research, Rafalski (2015), in the area of specific targeted feedback in a large urban school district in Florida.

While Rafalski's (2015) study examined if teachers were receiving specific targeted feedback and the effects on VAM scores, the research in this study sought to examine the contents of specific targeted feedback and how it affected student achievement outcomes especially level 1 and level 2 learners.

This researcher's study revealed that there was limited specific targeted feedback aimed at increasing the achievement of level 1 and level 2 students. This study also exhibited very limited correlations between different Marzano elements, as well categories and charges of feedback in relationship to student achievement level output on FCAT 2.0. There were some significant correlations found after calculating Pearson's r for elements scored and student achievement output; however, further data analyses showed the limited number of feedback comment data for specific elements may have affected the results. The results for one category of feedback (procedural) were inconsistent among assessments, but it was the only type of feedback with significant correlations. The fact that there were no significant correlations between the charge of feedback and student achievement was a puzzling find as information from the theoretical framework and review of literature suggested individuals might better respond to situations involving success and may have difficulties coping in situations involving failure.

The theoretical framework of Social Cognitive Theory (Bandura, 1977) indicated that an individual's self-efficacy might increase or decrease dependent upon the amount of successes and failures experienced. If observation comments consisted of information a teacher believed he or she was incapable of performing, then no amount of feedback

would help without some sort of successful experience, which could have a negative impact on student achievement. The goal of observational feedback was to help teachers reflect on their instructional practices and make changes needed to increase student achievement; however, if teachers equate observational feedback with failure, then that could lower self-efficacy and create a cycle of resistance. It may be important to determine a successful way to deliver feedback to teachers in order to support the successes of all stakeholders. It may also be vital to discover an effective method to optimize the current observational system in order to benefit teachers and students. While an observation framework might be helpful in guiding teachers to using best pedagogical practices, observers should be mindful that while they are observing teachers they are careful not to let the observation tool distract them from what is actually occurring in the classroom.

APPENDIX A
UNIVERSITY OF CENTRAL FLORIDA INSTITUTIONAL REVIEW
BOARD APPROVAL OF RESEARCH



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901, 407-882-2012 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

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

To : **Rachel Haynes**

Date : **July 21, 2015**

Dear Researcher:

On 07/21/2015 the IRB determined that the following proposed activity is not human research as defined by DHHS regulations at 45 CFR 46 or FDA regulations at 21 CFR 50/56:

Type of Review: Not Human Research Determination
Project Title: CONTENT AND EFFECTS OF SPECIFIC TARGETED
FEEDBACK FROM TEACHER OBSERVATIONS ON
STUDENT ACHIEVEMENT IN A LARGE URBAN
SCHOOL DISTRICT
Investigator: Rachel Haynes
IRB ID: SBE-15-11466
Funding Agency:
Grant Title:
Research ID: N/A

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

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

A handwritten signature in black ink that reads "Joanne Muratori".

Signature applied by Joanne Muratori on 07/21/2015 01:08:04 PM EDT

IRB manager

APPENDIX B
LARGE URBAN SCHOOL DISTRICT INSTITUTIONAL REVIEW
BOARD APPROVAL OF RESEARCH



Orange County Public Schools

445 West Amelia Street • Orlando, FL 32801-1129 • Phone 407.317.3200 • www.ocps.net

chievement in a Large Urban School District
Rachel Haynes

Dr. Barbara Murray

University of Central Florida

10/19/16

9/19/16

APPENDIX C
SAMPLE RUBRIC FROM RAFALSKI STUDY

No Feedback Provided (Level 1)	Unrelated Feedback or General Statement Provided (Level 2)	Recount of Classroom Events (Level 3)- Justification for rating	General Affirmation Statement (Level 4)	Reflective Question (Level 5)	Standardized Feedback Provided (Level 6)	Specific Targeted Feedback Provided (Level 7)
	2-The message is Unintelligible	3-Recap has several different components (sometimes statement of percent of students being monitored or desired effect). Sometimes actually gives examples of what is wrong with no suggestion. You did this or that, teacher did this or that, students did this or that, I observed this or that...	4-General praise. Good job, great job, excellent job, I liked, I loved, WOW!	5- Asks the teacher a question.	6- Examples: How might you adapt and create new strategies for chunking content into digestible bites that address unique student needs and situations? How might you expand your monitoring to involve more students? What are you learning about your students as you adapt and create new strategies? In addition to monitoring students by the use of choral responses, how else can you monitor students when chunking information?	7-Language like: 1. Reference to Resource Library or Reflective Teacher 2. Maybe try.... Or You might want to try.... 3. Consider.... 4. Recommendation.... Or I would recommend.... 5. Suggestion.... Or I suggest.... 6. It might be a good idea... 7. You should.... 8. This would have been good or great if... 9. To move to a higher level, do this_____ 10. Think about.... 11. I want you to.... 12. Next time.... 13. Always.... 14. Be sure to... or Make sure you.... 15. Doing this_____ would have been more effective. 16. You need to.... 17. Continue to... 18. Remember this_____ 19. Do this_____ 20. This_____is a good strategy.

No Feedback Provided (Level 1)	Unrelated Feedback or General Statement Provided (Level 2)	Recount of Classroom Events (Level 3)- Justification for rating	General Affirmation Statement (Level 4)	Reflective Question (Level 5)	Standardized Feedback Provided (Level 6)	Specific Targeted Feedback Provided (Level 7)
						21. I would like to see.... 22. Coaching idea...

APPENDIX D
RUBRIC FOR SPECIFIC TARGETED FEEDBACK CATEGORIES

Coaching	Professional Development	Procedural	Pedagogy	Content
The observer gives an explicit coaching tip or refers the teacher to other professionals for support.	The observer suggests specific professional development to the teacher to improve instruction.	The feedback is related to classroom rules, procedures, and teacher rapport.	The feedback is related to specific teaching practices or strategies.	The feedback is related to the specific content or subject-area being taught.
Gives tips on how to use other strategies	References the teacher to specific pages in the Marzano or other teaching book	Refers to rules, procedures, classroom management, or classroom layout	Gives information about teaching	Comments are directly related to the subject matter or content of the class
May say “coaching tip,” “recommendation,” or “suggestion”	Refers the teacher to another teacher, coach, or department head for help	Asks teacher to keep students on task	Advice to simply continue using the Marzano strategy	
Asks reflective questions but gives examples of what the teacher could do	Suggests videotaping or recording him or herself	Asks teacher to make sure students raise their hands or rotate groups a specific way	Telling teachers to make sure students understand something without coaching as to how	
Gives alternate strategies to use	Suggestion to observe other teachers	Gives suggestions for procedures for talking	Discusses monitoring students (as coaching advice)	
Does not only tell the teacher what to do but possible ways to use or practice a strategy		References behavior plans such as CHAMPS	Teaching styles, instructional methods	
Gives instruction or advice				

Note. While coaching tips and pedagogy are very closely related, there needed to be a separate code because not all feedback related to pedagogy consisted of coaching tips. Coaching tips are more specific, not only does the administrator want to see the improvement, he or she gives the teacher advice on *how* or specific ideas.

Note. If feedback can be coded as more than one theme, the predominant message is coded.

APPENDIX E
CHARGE RUBRIC

Positive	Negative	Neutral
Of or relating to good or desirable qualities	Of or relating to qualities that are undesirable	Having no strong good or undesirable qualities
Students are engaged	Focuses on what the teacher did not do which resulted in an issue with the lesson	Statements about what teachers should do
The teacher is making progress on the element	Focuses on what the students were not doing	Statements on how teachers can move up a level
The administrator liked what he or she saw	Focuses on what the teacher did not do	There is no praise or criticism
The administrator uses a positively charged praise like “good job,” “I liked,” “excellent,” etc.	A teacher may have tried the strategy but the feedback highlights the failure The teacher did not monitor	

Note. If feedback included a type of praise and a corrective action because students were not monitored or all were on task, the feedback was coded as positive

Note. If feedback begins with a type of praise but ends with a criticism, the feedback was coded as negative

APPENDIX F
CHARGE AND CATEGORY EXAMPLE MATRIX

	Coaching Tip	Procedural	Pedagogy	Professional Development	Content
Positive	<p>“Good use of grouping strategies to help students interact with the new knowledge. In the future, you may want to provide students with group responsibilities to help with the collaborative process. This will add more structure to your groups and allow for deeper levels of discussion. This will also help you monitor at all students are participating in the collaborative process.”</p>	<p>“Good job of having students read the learning goal together out loud at the beginning of class. Went over the scale and had students identify their level on it before reading Achieve 3000 article and then again after finishing the whole group activity. Good job of having the students tell you why they feel they are at the scale level they were at after the whole group instruction for validation. Need to make sure it is done daily so students recognize the consistency and its importance to the success of their learning.”</p>	<p>“[HIDDEN]'s use of learning goal and scale has improved drastically from the beginning of this year. To move to innovating, the desired effect must be evident in 100% of the students.”</p>	<p>“Nice job with dealing with lack of adherence to rules and procedures. [HIDDEN] signaled the group twice due to loud voices. She explained that their would be a consequence if students continued to be loud during their group work. Students were warned again about loud voices and a frowny face was added to the board. Other strategies that can be used can be found in the book, Becoming a Reflective Teacher pgs. 160-162. To move from applying to innovating, how can you adapt and create new strategies for unique student needs and situations in order for the desired effect to be</p>	<p>“At the start of the lesson [HIDDEN] reviewed what they have been working on, specifically what they worked on yesterday. She reviewed specific examples, an earring box, and each of the steps with the students to find the volume of a prism. Great examples with the pictures within the PP. I like that you are giving the students step by step directions on how to find the volume of a prism. Have you taught the academic vocabulary that goes along with this lesson? Length, width, height, etc. in your step by step directions you should include the correct terminology to go along with the lesson.</p>

	Coaching Tip	Procedural	Pedagogy	Professional Development	Content
				evident in all students?"	Consider having the students complete a quick problem as I stated in element 1. Some of the students were following the review and some were not. Perhaps given a short example to identify the length, width and height, even if they didn't come to the conclusion of identifying volume. At least you could see who could find the three measurements needed to solve for volume."
Neutral	"One suggestion would to walk around room as students practice You did walk across the middle of the rows while student was putting information on the board. A suggestion would be to look at	"Continue to have students describe what procedure or rule they need to follow to help them internalized it and minimize distractions."	"Continue to focus on helping students deepen their knowledge by examining similarities and differences."	"Review pages 160-162 from Becoming a Reflective teacher to apply consequences for not following rules and procedures consistently and fairly."	"[HIDDEN] shows the class one problem from last nights homework, with which he found that most students had trouble. He guides students through analyzing the plotted line graph, scales, and

	Coaching Tip	Procedural	Pedagogy	Professional Development	Content
	the work students produced.”				data table provided. Most students chose "multiply by three" as the answer, when the correct answer was "multiply by 1/3". Students pay attention to this and some "oh"s can be heard around the room. One suggestion, take time to explain why 1/3 is different - your lower performing students could benefit from a brief review of why the data table must be read as "x times 1/3" rather than "y times 3".”
Negative	“You were focused on the presenter, but did not monitor the engagement of the other students. Suggestion, have students ask the presenter questions based on their presentation. Each	“As teacher was reviewing the problems on the board, students continued to talk with little or no redirection. The teacher asked two students to stop talking but they	“You did not celebrate any successes of gains towards the goal. Check progress and celebrate. This lets students know that they can be successful and also be acknowledged for	“Questions posed did not allow for elaboration. To move to applying, include questions that require inferencing to allow students to elaborate. Page 113 in reflective	“You will be able to identify the main idea of a story or article. In addition to providing a clearly stated learning goal accompanied by a scale or rubric that describes levels of performance, how

Coaching Tip	Procedural	Pedagogy	Professional Development	Content
<p>student must ask at least 1 or 2 questions and they never know which presenter they'll have to ask questions for, this is a good trick to force students to pay attention. Most students were completely tuned out.”</p>	<p>continued. One suggestion would be to have a cue for students to know when all talking stops. After 12 minutes the teacher says "I hear a lot of talking back here and it needs to stop". However, talking did continue without a consequence or reminder of rules.”</p>	<p>their hard work.”</p>	<p>educator discusses using default questions and reasoned inference questions. These strategies will help you move to developing and applying.”</p>	<p>can you monitor students understanding of the learning goal and the levels of performance? Recommendation: reference the learning goal and link it to everything you do frequently during the lesson. The students should understand that this is the whole point of the lesson. You did a quick review of the main idea at the end, which was not sufficient to ensure they understood. You also very briefly touched on main idea during your small group, however you could have done this as they read each paragraph to really cement and assess comprehension and mastery.”</p>

APPENDIX G
EXAMPLES OF IMPROPER FEEDBACK COMMENTS

Comment

“During guided practice several students asked questions and teacher referred them to a partner or asked if they had asked someone else. One suggestion would be to assist the student in the questions since this appeared to be a new task and the partner may have the wrong answer and/or not understand it either. The class **if** very fragmented with teacher talking and students not listening. More than 8 students expressed that they did not understand what and/or why the answer **way** chosen. Students did not want to ask questions. The student that was absent on the previous day was told to check with a neighbor.”

“With 29 students in a co-taught class organizing students to interact with new knowledge in groups would allow for better monitoring for the desired effects. One **suggestions** would be to also allow the co-teacher to provide instruction rather than as a teacher aid. This would then allow you to provide some direct assistants to struggling students.”

“[HIDDEN] has chunked the topics for students to compare and contrast on the **vin** diagram. Each small group discusses just that section to compare and contrast and then they share with the other groups. This was more effective **then** saying compare and contrast 120+ pages of the book we have read so far. Suggestion: in the rotation to share, include directions to ask the other partner, do you remember anything we may have **forgotten**. This helps gets kids thinking about the whole book.”

“Having students come to the overhead was a good strategy however, instead of telling the student the answer you **could of** asked him why and how he got his answer and ask others if they got the same answer or a different one and if different they could come up and write what they put and have the class **examin** both answers.”

“You called on some students **to to** answer questions, but **no half** the students. This is a great opportunity to use the whiteboards. Require students to explain their answers to the group, you circulate and monitor the reasons.”

“Pace of the class is too slow and monotone. **Energy level and engagement not consistent for majority of class**. Whole group instruction of Achieve 3000 took too long and transition to other rotations needs to be sooner so more time is used in the other stations. Speeding up the lesson with more enthusiasm will also keep students more **focussed** throughout the lesson and not rely on the answers from those engaged. I also would have the computers already on and at the sign-in screen to speed up that transition. Set the timer up so it notifies you when the transition needs to occur from the whole group instruction to the rotations. Watch the time so you have time to wrap up class and do a final scale check before bell rings.”

“feedback:

Nice job of using the scale! You understand the motivational purpose and the idea of validating the level they think they are at.
YOU GOT THIS!!!

Make **id** a routine event in everything. Also secure the CHAMPS expectations throughout **th eday** :)”

“[HIDDEN] practices the teach **like** a **CHampion** techniques to keep engagement high. Remember to recognize individual students”

“[HIDDEN] speaks to the students with an energetic **voise**, he communicates his excitement to see his kids **be able** to improve their skill after today's lesson. By speaking to his class in this manner, I can see he understands the point of having a goal and a scale for motivation and to increase competence of a **skill.My** only suggestion is to talk less, you set up rigorous situations and as soon as you begin to talk, there is a **tendence** for the students to move out of a concentration on skill or problem solving. Observe and monitor by quietly walking

Comment

around and look at their work. We will discuss this at the post conference. Your students are so attracted to your style of talking to them. Your coaching words and tone are extremely motivating and provide **hpe** for the kids who lack a drive for success.”

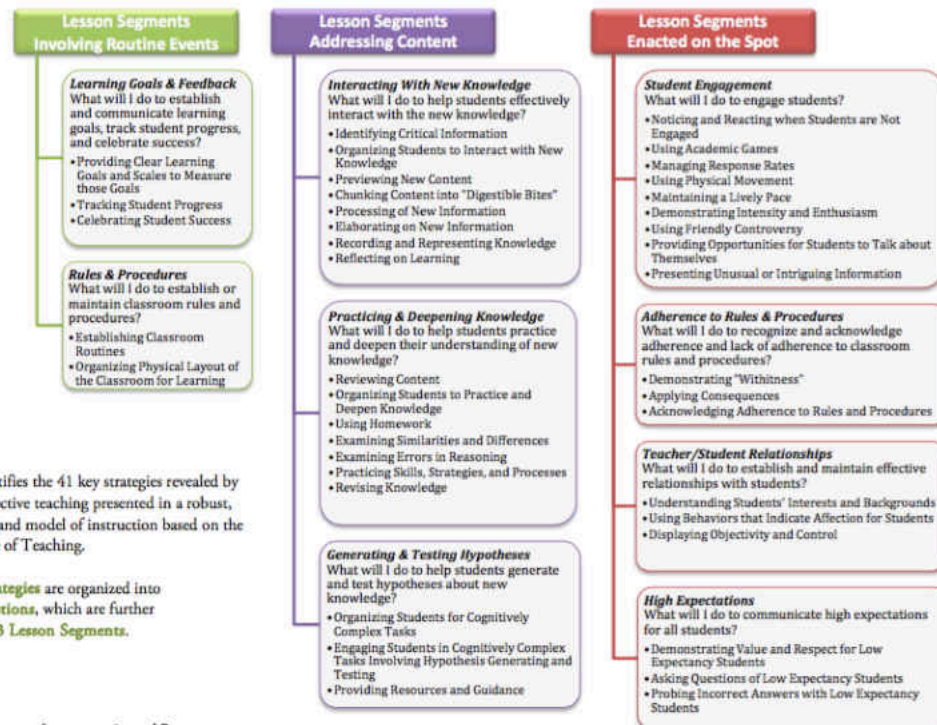
“The mini-lesson is suppose to be 10 minutes maximum, but it lasted for approximately 35 to 40 minutes which is entirely too long. Make sure you pace **yoourself** so that your mini-lesson doesn't become the whole group lesson and then your unable to get to your lesson that you planned.”

“Your student desks are arranged in two tables of long rows with one student by himself in the front of the classroom. You have clear traffic patterns and small group areas. However, you do not have your common board **are** filled out so that you are missing essential questions and learning goals.”

Note. These comments are examples of feedback that contains misspelled words, incorrect sentence syntax and structure, and improper punctuation. The mistakes in the comments are bolded and underlined.

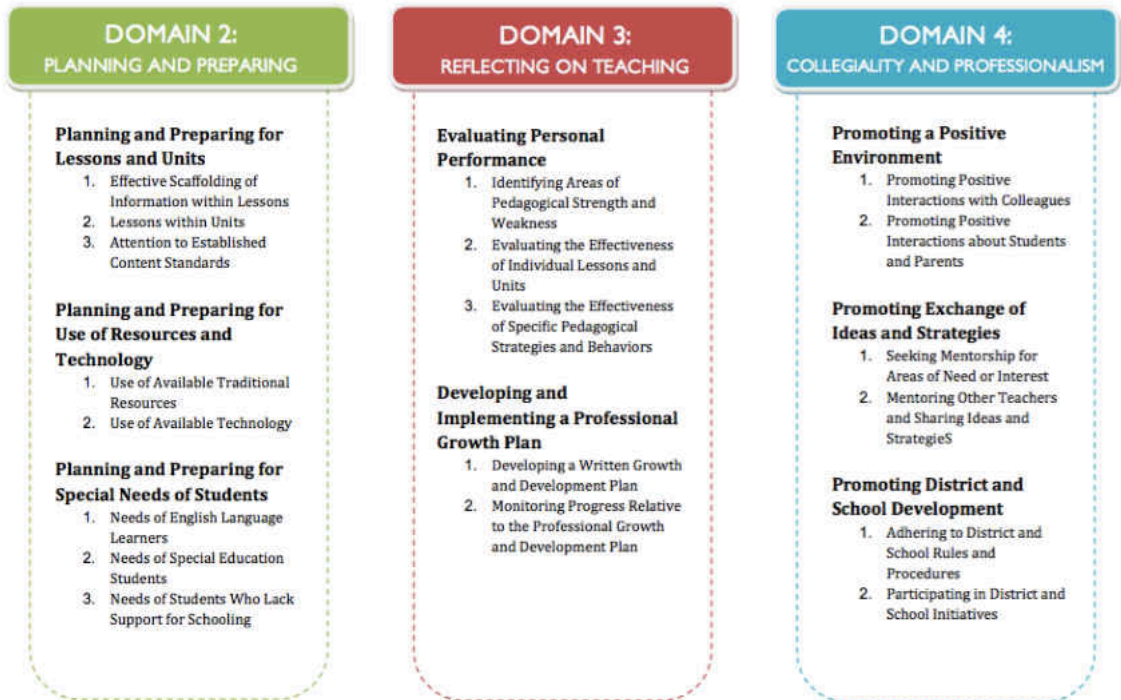
APPENDIX H
MARZANO LEARNING MAP

DOMAIN 1: CLASSROOM STRATEGIES AND BEHAVIORS



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

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



REFERENCES

- Accountability. (n.d.). In dictionary.com. Retrieved from <http://dictionary.reference.com/browse/accountability>.
- Anast-May, I., Penick, D., Schroyer, R., & Howell, A. (2011). Teacher conferencing and feedback: Necessary but missing.
- Assessment and accountability briefing book: FCAT, school accountability, teacher certification tests. (2007). Retrieved from Florida Department of Education website: <http://fcat.fldoe.org/pdf/BriefingBook07web.pdf>.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Brimley, Jr., V, Verstegen, D.A., Garfield, R.R. (2016). Financing education in a climate of change. Pearson.
- Bryant, J., & Zillman, D. (Eds.) (2002). Media effects: Advances in theory and research (2nd ed., pp. 121-153). Hillsdale, NJ: Lawrence Erlbaum.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researchers*, Vol. 33, No. 8, pp. 3-15. Retrieved from....
- Cameron, S., Mulholland, J, & Branson, C. (2013). Professional learning in the lives of teachers: Towards a new framework for conceptualising teacher learning. *Asia-Pacific Journal of Teacher Education*. 41:4, 377-397, DOI: 10.1080/1359866X.2013.838620.
- Darling-Hammond, L. (1998). Teacher learning that supports student learning. *Educational Leadership*, 5. 6-11.

- Darling-Hammond, L., Chung Wei, R., Andree, A., Richardson, N., & Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the united states and abroad. National Staff Development Council. Retrieved from <http://learningforward.org/docs/pdf/nsdcstudy2009.pdf>.
- Fact sheet: The race to the top. (2009, November 04). Retrieved from www.whitehouse.gov/the-press-office/fact-sheet-race-top.
- Florida Department of Education. (2013). Accountability results. Retrieved from <http://www.fldoe.org/accountability/assessments/k-12-student-assessment/results/2013.stml#FCAT2>.
- Florida Department of Education. (2011, April). American Institute for Research. Retrieved from www.fldoe.org/committees/pdf/sgPresentation.pdf.
- Florida Department of Education. (2014). Fcat 2.0 and florida eoc assessments achievement levels. Retrieved from <http://www.fldoe.org/core/fileparse.php/3/urlt/achlevel.pdf>.
- Florida Department of Education. (2014). FCAT 2.0: Florida statewide assessments 2014 technical report. Retrieved from https://accountabaloney.files.wordpress.com/2016/05/fcat_2-0_fl_eoc_2014_technical_report_final_distribute.pdf.
- Florida Department of Education. (2015). Florida educator certification renewal requirements. Retrieved from <http://www.fldoe.org/teaching/certification/renewal-requirements#AfterJuly>.

Florida Department of Education. (n.d.). Florida's value added model: Overview of the model to measure student learning growth on fcat as developed by the student growth implementation committee. Retrieved from www.fldoe.org/committees/pdf/PresentationValue-addedModel.pdf.

Florida Department of Education. (n.d.). Performance evaluation. Retrieved from <http://www.fldoe.org/teaching/performance-evaluation>.

Florida Department of Education. (2015). Professional development standards. Retrieved from <http://www.fldoe.org/teaching/professional-dev/professional-dev-stands.shtml>.

Florida Department of Education. (n.d.). Recommendations of the florida student growth implementation committee: Background and summary. Retrieved from www.fldoe.org/committees/.../Value-Added-Model-White-Paper.doc.

Florida Department of Education. (2012). Teacher evaluation systems alignment: The florida educator accomplished practices (FEAPs), dr. Robert marzano and charlotte danielson. Retrieved from <http://www.fldoe.org/core/fileparse.php/7503/urlt/0071814-tesa-feaps-marzanodanielson.pdf>.

Florida Department of Education. (2013). Understanding fcat 2.0 reports. Retrieved from <http://fcats.fldoe.org/mediapacket/2013/pdf/2013UFR.pdf>.

Florida Department of Education (2014). 2013-2014 fcat 2.0 reading, mathematics, science, and writing fact sheet. Retrieved from

<http://www.fldoe.org/core/fileparse.php/5668/urlt/0101323-201314factsheet20.pdf>.

Florida State Statute § 1008.31 (a). Assessment and accountability. (2012).

Florida State Statute § 1012.34(1)(a). Personnel evaluation procedures and criteria. Evaluation system approval and reporting (2012).

Florida State Statute § 1012.34(3)(1). Personnel evaluation procedures and criteria. Evaluation procedures and criteria (2012).

Florida State Statute § 1012.34. Personnel evaluation procedures and criteria. Evaluation procedures and criteria (2013).

Florida State Statute § 1012.98(1). School community professional development act (2012). Retrieved from

<http://www.fldoe.org/core/fileparse.php/5668/urlt/0101323-201314factsheet20.pdf>.

Frase, L. E. (1992). Constructive feedback on teaching is missing. *Education*, 113(176-181). Retrieved from

<http://eds.b.ebscohost.com.ezproxy.net.ucf.edu/ehost/pdfviewer/pdfviewer?sid=68c710f9-4b8a-4a1d-8f3b-982b85107f94%40sessionmgr198&vid=9&hid=122>

Gabriel, J. G. (2005). How to thrive as a teacher leader. Alexandria, VA: Association for Supervision and Curriculum Development.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review Of Educational Research*, 77(1), 81-112. doi: 10.3102/003465430298487.
- Hill, D. (2000, May). He's got your number. *Teacher magazine*, 11(8), 42-47. Retrieved from <http://ezproxy.net.ucf.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=507693974&site=ehost-live>.
- Hill, H. C., & Grossman, P. (2013). Learning from Teacher Observations: Challenges and Opportunities Posed by New Teacher Evaluation Systems. *Harvard Educational Review*, 83(2), 371-384. Retrieved from <http://hepgjournals.org.ezproxy.net.ucf.edu/doi/pdf/10.17763/haer.83.2.d11511403715u376>.
- Hoekstra, A., Brekelmans, M., Beijaard, D., & Korthagen, F. (2009). Experienced teachers' informal learning: Learning activities and changes in behavior and cognition. *Teaching & Teacher Education*, 25(5), 663-673. doi: 10.1016/j.tate.2008.12.007.
- Kelting, T., Jenkins, J. M., & Gaudreault, K. L. (2014). I Could Really Focus on the Students. *JOPERD: The Journal Of Physical Education, Recreation & Dance*, 85(8), 32-37. doi: 10.1080/07303084.2014.946192.
- Kwakman, K. (2003). Factors affecting teachers' participation in professional learning activities. *Teaching & Teacher Education*, 19(2), 149-170. doi: 10.1016/S0742-051X(02)00101-4.

- Learning Sciences International. 2010. Developing policies and procedures: Webinar 1: calculating and weighting the instructional practice score for the annual evaluation. Retrieved from http://www.marzanoevaluation.com/files/DPP_Day_1_Calculating_and_Weighting_WEBINAR_OK_21020906.pdf.
- 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 Press.
- Marzano, R., J. (2011). Marzano art and science of teaching: Teacher evaluation model. Retrieved from http://www.marzanoevaluation.com/files/Day_1/LearningMap_4Domains.pdf.
- Marzano, R. J., (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.
- Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State Of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25:2, 231-256, DOI: 10.1080/09243453.2014.885451.
- Nolan, J. F. & Hoover, L. A. (2011). Teacher supervision and evaluation: Theory into practice. John Wiley & Sons, Inc.

- Orange County Public Schools. (2012). Instructional personnel education system procedures manual: 2012-2013. Retrieved from www.nctq.org/docs/Evaluation_Manual_2012-13_Final.pdf.
- Orange County Public Schools. (2011). Teacher evaluation frequently asked questions. Retrieved from <https://www.ocps.net/cs/pds/assessment/Additional%20Information/Teacher%20Evaluation%20-%20Frequently%20Asked%20Questions.pdf>.
- Orange County Classroom Teachers Association & The School Board of Orange County, Florida. (2013-2014). Contract between The School Board of Orange County, Florida And The Orange County Classroom Teachers Association. Retrieved from [https://www.ocps.net/es/legislative/laborrelations/Documents/2014-2015/CTA%20Contract%2013-14%20\(3\).pdf](https://www.ocps.net/es/legislative/laborrelations/Documents/2014-2015/CTA%20Contract%2013-14%20(3).pdf).
- Owens, R. G, & Valesky, T. C. (2011). Organizational behavior in education: Leadership and school reform. Upper Saddle River, NJ: Pearson Education, Inc.
- Owston, R. (2008). Models and methods for evaluation. *Handbook of Research on Educational Communications and Technology*, 605-617. Retrieved from http://www.aect.org/edtech/edition3/ER5849x_C045.fm.pdf.
- Quinn, A. E. (2014). Looking at the Bigger Picture with Dr. Robert Marzano: Teacher Evaluation and Development for Improved Student Learning. *Delta Kappa Gamma Bulletin*, 81(1), 12-18. Retrieved from <http://eds.b.ebscohost.com.ezproxy.net.ucf.edu/ehost/pdfviewer/pdfviewer?sid=712d156e-3c62-4f2a-bf03-14a5d97a7140%40sessionmgr110&vid=14&hid=122>.

- Rafalski, S., H. (2015). Policy implications of a teacher evaluation system: The relationship of classroom observations, levels of feedback, and student achievement outcomes.
- Range, B., Duncan, H., & Hvidston, D. (2013). How Faculty Supervise and Mentor Pre-Service Teachers: Implications for Principal Supervision of Novice Teachers. *International Journal Of Educational Leadership Preparation*, 8(2), 43-58. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1016279.pdf>.
- Range, B. G., Young, S., & Hvidston, D. (2012). Teacher perceptions about observation conferences: What do teachers think about their formative supervision in one US school district? *School leadership & management: Formerly school organisation*, 33:1, 61-77, doi: 10.1080/13632434.2012.724670.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Books.
- Reeves, D. B. (2009). *Leading change in your school: How to conquer myths, build commitment, and get results*. Alexandria, VA: ASCD.
- Steinberg, W. J. (2011). *Statistics alive*. Thousand Oaks, CA: Sage Publications, Inc.
- Tang, S. F., & Chow, A. K. (2007). Communicating feedback in teaching practice supervision in a learning-oriented field experience assessment framework. *Teaching & Teacher Education*, 23(7), 1066-1085. doi: 10.1016/j.tate.2006.07.013.
- Teacher evaluation frequently asked questions: 2012-2013. (n.d.). Retrieved from Orange County Public Schools website:

www.ocps.net/es/hr/PDS/assessment/Additional%20Information/Teacher%20Evaluation%20-%20Frequently%20Asked%20Questions.pdf.

Tuytens, M., & Devos, G. (2014). How to Activate Teachers through Teacher Evaluation?. *School Effectiveness And School Improvement*, 25(4), 509-530. doi: 10.1080/09243453.2013.842601.

U.S. Department of Education. (2003, August 23). Fact sheet on the major provisions of the conference report to h.r. 1, the no child left behind act. Retrieved from www2.ed.gov/print/nclb/overview/intro/factsheet.html.

U.S. Department of Education. (2002). No child left behind: A desktop reference. Retrieved from <https://www2.ed.gov/admins/lead/account/nclbreference/reference.pdf>.

U.S. Department of Education. (2009). Race to the top executive summary. Retrieved from <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.

U.S. Department of Education. (2010, May 27). Race to the top program: Guidance and frequently asked questions. Retrieved from www2.ed.gov/programs/racetothetop/faq.pdf.

Wade, R., K. (1984). What makes a difference in inservice teacher education? A meta-analysis of research. *Educational Leadership*