

---

Electronic Theses and Dissertations, 2004-2019

---

2013

## An Analysis Of The School Board Of Brevard County Instructional Personnel Performance Appraisal Instrument And The Validity Of Its Components

Carol Mela  
*University of Central Florida*



Part of the [Educational Leadership Commons](#)

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

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

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

---

### STARS Citation

Mela, Carol, "An Analysis Of The School Board Of Brevard County Instructional Personnel Performance Appraisal Instrument And The Validity Of Its Components" (2013). *Electronic Theses and Dissertations, 2004-2019*. 2560.

<https://stars.library.ucf.edu/etd/2560>



AN ANALYSIS OF THE SCHOOL BOARD OF BREVARD COUNTY  
INSTRUCTIONAL PERSONNEL PERFORMANCE APPRAISAL  
INSTRUMENT AND THE VALIDITY OF ITS COMPONENTS

by

CAROL ANN MELA  
B.S. Bowling Green State University, 1982  
M.S. Ashland College, 1988  
Ed.S. Nova Southeastern University, 1998

A dissertation submitted in partial fulfillment of the requirements  
for the degree of Doctor of Education  
in the Department of Educational Leadership  
in the College of Education  
at the University of Central Florida  
Orlando, Florida

Spring Term  
2013

Major Professor: Barbara Murray

© 2013 Carol Mela

## ABSTRACT

This study examined the relationship between the professional practices components of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument and student achievement as measured by a teacher's value-added measurement score. A Pearson Product-Moment Correlation was the statistical test used to analyze the data. The population included Brevard Public School instructional personnel assigned to Grades 4-10 who taught reading and/or mathematics measured by the Florida Comprehensive Assessment Test during the 2011-2012 school year and who received teacher aggregated value-added measurement scores.

Findings indicated that there was a small to moderate statistically significant, positive relationship between all eight professional practices component variables and value-added measurement scores. Correlation coefficients ranged from .089 for collaborative inquiry to .218 for quality of instruction. All of the components combined had a correlation of .231, confirming the strength of multiple evaluation measures.

Recommendations were provided for future research aimed at further data analysis in Brevard Public Schools as well as other school districts in order to identify the combination of evaluation components that most accurately reflect teaching effectiveness resulting in student learning as well as to pinpoint weaknesses upon which additional training could be based and the fidelity of implementation improved.

## ACKNOWLEDGMENTS

It is not without the support of several people that one successfully completes a doctoral program, and it was no different for me. I cannot express the gratitude that I feel toward the following individuals who contributed to my success:

- My mother and father who taught me to always do my best at everything I do and to continually learn and develop by setting goals and pursuing them.
- My husband whose love, support, and encouragement during my doctoral program provided me with the ability to place my studies center stage.
- My daughter who supported me by listening to me talk about my “college life” while she shared hers.
- My University Of Central Florida professors who challenged my thinking and broadened my knowledge base.
- My advisor and dissertation chair, Dr. Barbara Murray, along with my dissertation committee comprised of Dr. Debbie Hahs-Vaughn, Dr. Kenneth Murray, and Dr. Rosemarye Taylor who guided me throughout my dissertation and provided me with the support I needed.
- My assistant principal, Mrs. Jennifer Brockwell, who took on many additional responsibilities to assist with my caseload as well as served as a sounding board for my dissertation.
- My school staff that were patient and understanding with me throughout my entire doctoral program.
- The Brevard Public School System and all of the district employees who assisted with getting data to me and answering my many questions.

# TABLE OF CONTENTS

LIST OF FIGURES .....	xi
LIST OF TABLES .....	xv
CHAPTER I INTRODUCTION.....	1
Instrument Development.....	5
Purpose of the Study .....	5
Statement of the Problem.....	6
Research Questions and Hypotheses .....	7
Delimitations.....	9
Limitations .....	10
Definitions of Terms .....	10
Theoretical Framework.....	12
Summary.....	16
Organization of the Study .....	17
CHAPTER II LITERATURE REVIEW .....	18
Teacher Quality.....	19
Historical Perspective .....	22
Observation.....	28
Los Angeles Elementary Charter School.....	29
Washoe County, Nevada School District.....	31
Arizona.....	31
Midsize Western United States School District.....	32
Cincinnati Public School System.....	33

A Measure of Effective Teaching Project.....	35
Northeastern Florida Schools.....	36
Texas Independent School District .....	37
Summary .....	38
Individualized Growth Plans .....	38
Definition .....	39
Underlying Assumptions .....	40
Impact on Student Achievement.....	42
Summary .....	45
Collaboration and Mutual Accountability .....	45
Definition .....	45
Teacher Outcomes.....	46
Student Outcomes .....	47
Summary .....	51
Value-Added Assessment.....	51
Definition .....	51
Models.....	52
Reliability.....	54
Validity .....	56
Outcome Measures.....	57
Florida’s Model.....	59
Summary .....	60
CHAPTER III METHODOLOGY .....	61

Research Design .....	61
Population .....	62
Methods of Data Collection .....	63
Instrumentation .....	64
Reliability .....	65
Classroom Observation Instrument .....	67
Observational Rubrics .....	69
BPS Instructional Personnel Evaluation Instrument .....	73
Professional Growth Plan Template .....	76
Individual Professional Growth Plan Development Rubrics .....	79
Professional Growth Plan Development Scoring and Feedback Form .....	80
Professional Growth Plan Development Administrator Final Scoring and Feedback Form .....	81
Mid-Year Conference Form .....	82
Individual Professional Growth Plan Implementation Rubrics .....	82
Professional Growth Plan Implementation Scoring and Feedback Form .....	84
Collaboration & Mutual Accountability Team Scoring Form .....	84
Value-Added Growth Measure .....	85
Data Analysis .....	87
Summary .....	90
CHAPTER IV ANALYSIS OF DATA .....	91
Descriptive Statistics .....	91
Population .....	91



Findings .....	93
Research Question and Hypothesis #1 .....	93
Research Question and Hypothesis #2.....	101
Research Question and Hypothesis #3.....	105
Research Question and Hypothesis #4.....	110
Research Question and Hypothesis #5.....	114
Research Question and Hypothesis #6.....	119
Research Question and Hypothesis #7.....	123
Research Question and Hypothesis #8.....	127
Summary.....	132
CHAPTER V SUMMARY, DISCUSSION, AND CONCLUSIONS.....	133
Summary of the Study .....	133
Discussion of the Findings.....	136
Research Question One.....	137
Research Question Two .....	140
Research Question Three .....	142
Research Question Four .....	143
Research Question Five, Six, and Seven.....	145
Research Question Eight.....	147
Implications for Practice.....	148
Further Recommendations for Brevard Public School District .....	151
Recommendations for Future Research.....	153
Conclusions.....	156

APPENDIX A PERMISSION TO REPRODUCE – DR. JAMES STRONGE .....	157
APPENDIX B OVERVIEW OF SEARCH PARAMATERS .....	159
APPENDIX C PERMISSION TO REPRODUCE – DR. LAURA GOE.....	162
APPENDIX D PERMISSION TO REPRODUCE – DR. DEBRA PACE.....	164
APPENDIX E COURSE LIST .....	167
APPENDIX F EVALUATION SUMMARY CALCULATION FORM .....	170
APPENDIX G IRB REVIEW AND BREVARD PUBLIC SCHOOLS RESEARCH APPROVAL .....	173
APPENDIX H CLASSROOM OBSERVATION INSTRUMENT .....	176
APPENDIX I OBSERVATION RUBRICS .....	181
APPENDIX J BPS INSTRUCTIONAL PERSONNEL EVALUATION INSTRUMENT .....	195
APPENDIX K INSTRUCTIONAL PROFESSIONAL GROWTH PLAN TEMPLATE .....	200
APPENDIX L INDIVIDUAL PROFESSIONAL GROWTH PLAN DEVELOPMENT RUBRICS .....	202
APPENDIX M PROFESSIONAL GROWTH PLAN DEVELOPMENT SCORING AND FEEDBACK FORM .....	205
APPENDIX N PROFESSIONAL GROWTH PLAN DEVELOPMENT ADMINISTRATOR FINAL SCORING AND FEEDBACK FORM.....	207
APPENDIX O MID-YEAR CONFERENCE FORM .....	209
APPENDIX P INDIVIDUAL PROFESSIONAL GROWTH PLAN IMPLEMENTATION RUBRICS.....	211

APPENDIX Q PROFESSIONAL GROWTH PLAN IMPLEMENTATION AND FEEDBACK FORM.....	213
APPENDIX R COLLABORATION & MUTUAL ACCOUNTABILITY TEAM SCORING FORM.....	215
APPENDIX S VALUE-ADDED MEASURES .....	217
REFERENCES .....	219

## LIST OF FIGURES

Figure 1: Goals and Roles Evaluation Model (Stronge & Tucker, 2003, p. 30). .....	13
Figure 2: Teacher Quality Framework (Goe, 2007, p. 9). .....	21
Figure 3: Performance Appraisal Model (Brevard Public Schools, 2011, p.17). .....	65
Figure 4: Continuous Improvement Cycle (Brevard Public Schools, 201, Appendix 36). .....	77
Figure 5: Teacher Aggregated VAM Histogram. ....	94
Figure 6: Normal Q-Q Plot of Teacher Aggregated VAM Scores. ....	94
Figure 7: Detrended Normal Q-Q Plot of Teacher Aggregated VAM Scores.....	95
Figure 8: Box Plot of Teacher Aggregated VAM Scores. ....	96
Figure 9: Professional Practices (Formative Evaluation Total) Histogram. ....	97
Figure 10: Normal Q-Q Plot of Professional Practices (Formative Evaluation Total) Ratings. ....	98
Figure 11: Detrended Q-Q Plot of Professional Practices (Formative Evaluation Total) Ratings. ....	98
Figure 12: Box Plot of Professional Practices (Formative Evaluation Total) Ratings. ...	99
Figure 13: Scatterplot of Professional Practices (Formative Evaluation Total) Ratings and Teacher Aggregated VAM Scores. ....	100
Figure 14: Instructional Practices (Component 1 - Cumulative Total) Histogram.....	101
Figure 15: Normal Q-Q Plot of Instructional Practices (Component 1 - Cumulative Total) Ratings.....	102

Figure 16: Detrended Normal Q-Q Plot of Instructional Practices (Component 1 - Cumulative Total) Ratings.....	103
Figure 17: Box Plot of Instructional Practices (Component 1 - Cumulative Total) Ratings. ....	103
Figure 18: Scatterplot of Instructional Practices (Component 1 - Cumulative Total) Ratings and Teacher Aggregated VAM Scores.....	104
Figure 19: Quality of Instruction Histogram. ....	106
Figure 20: Normal Q-Q Plot of Quality of Instruction Ratings. ....	107
Figure 21: Detrended Normal Q-Q Plot of Quality of Instruction Ratings. ....	107
Figure 22: Box Plot of Quality of Instruction Ratings.....	108
Figure 23: Scatterplot of Quality of Instruction Ratings and Teacher Aggregated VAM Scores.....	109
Figure 24: Professional Responsibility, Conduct and Relationships Ratings Histogram. ....	110
Figure 25: Normal Q-Q Plot of Professional Responsibility, Conduct and Relationships Ratings. ....	111
Figure 26: Detrended Normal Q-Q Plot of Professional Responsibility, Conduct and Relationships Ratings.....	112
Figure 27: Box Plot of Professional Responsibility, Conduct and Relationships Ratings. ....	112
Figure 28: Scatterplot of Professional Responsibility, Conduct and Relationships Ratings and Teacher Aggregated VAM Scores. ....	113

Figure 29: Professional Development (Component 2 - Cumulative Total) Ratings Histogram.....	115
Figure 30: Normal Q-Q Plot of Professional Development (Component 2 - Cumulative Total) Ratings.....	116
Figure 31: Detrended Normal Q-Q Plot of Professional Development (Component 2 - Cumulative Total) Ratings.....	116
Figure 32: Box Plot of Professional Development (Component 2 - Cumulative Total) Ratings. ....	117
Figure 33: Scatterplot of Professional Development (Component 2 - Cumulative Total) Ratings and Teacher Aggregated VAM Scores. ....	118
Figure 34: Action Research Development (Professional Growth Plan Development) Ratings Histogram. ....	119
Figure 35: Normal Q-Q Plot of Action Research Development (Professional Growth Plan Development) Ratings. ....	120
Figure 36: Detrended Normal Q-Q Plot of Action Research Development (Professional Growth Plan Development) Ratings. ....	121
Figure 37: Box Plot of Action Research Development (Professional Growth Plan Development) Ratings. ....	121
Figure 38: Scatterplot of Action Research Development (Professional Growth Plan Development) Ratings. ....	122
Figure 39: Implementation of Action Research (Professional Growth Plan Implementation) Ratings Histogram.....	124

Figure 40: Normal Q-Q Plot Implementation of Action Research (Professional Growth Plan Implementation) Ratings.....	125
Figure 41: Detrended Normal Q-Q Plot of Implementation of Action Research (Professional Growth Plan Implementation) Ratings. ....	125
Figure 42: Box Plot of Implementation of Action Research (Professional Growth Plan Implementation) Ratings.....	126
Figure 43: Scatterplot of Implementation of Action Research (Professional Growth Plan Implementation) Ratings.....	127
Figure 44: Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings Histogram.....	128
Figure 45: Normal Q-Q Plot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings. ....	129
Figure 46: Detrended Normal Q-Q Plot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings.....	130
Figure 47: Box Plot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings. ....	130
Figure 48: Scatterplot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings. ....	131

## LIST OF TABLES

Table 1: Professional Practices Correlation.....	139
--	-----



## CHAPTER I INTRODUCTION

To better prepare for students to compete in a global economy, student learning must be at a high level. To this end, teacher quality is a topic of high interest for educational reformers. A well-known study cited throughout the literature that focused on the relationship between teacher quality and student achievement was conducted by William L. Sanders and June C. Rivers in 1996 (Haycock & Hanushek, 2010). Sanders and Rivers (1996) concluded that the most prevalent factor influencing student achievement was teacher effect.

Numerous studies have been carried out using a value-added approach to estimate teacher effect. Haycock (1998), in her review of value-added research, found that teachers in the Boston Public Schools, when ranked according to assessment results, influenced learning quite differently. In fact, the top one-third of the teachers when compared to the bottom one-third produced six times the learning. Hanushek (1992), in studying low performing schools in New Jersey, likewise concluded that it was estimated that a student of a good teacher compared to a student of a bad teacher had an achievement growth difference of an equivalent of one grade level. Following many years of continued research on this topic, the results still confirmed the importance of teacher quality as stated by Arne Duncan, U. S. Secretary of Education:

Studies repeatedly document that the single biggest influence on student academic growth is the quality of the teacher standing in the front of the classroom - not socioeconomic status, not family background, but the quality of the teacher in front of that class. (Teachers College Columbia University, 2009, 15:31)

Improving the quality of teachers has been one of the major focuses of educational reforms over the last 50 years (Cohen-Vogel, 2005). Public Law 107-110, the No Child Left Behind Act of 2001 (NCLB), as a reauthorization of the Elementary and Secondary Education Act of 1965 (ESEA), was based on the idea that establishing high standards and measurable goals could increase student achievement (U.S. Department of Education, 2001). The Act required all students to be taught by highly qualified teachers who were defined according to their credentials and content knowledge.

On February 17, 2009, the American Recovery and Reinvestment Act (ARRA) was signed into law by President Obama (U.S. Department of Education, 2009). The ARRA made available \$4.35 billion for a competitive grant program referred to as “Race to the Top”. The single highest criterion for which points were awarded was improving principal and teacher effectiveness. This reform effort was designed to assist states in developing evaluation systems that linked teacher and student data.

On March 13, 2010, the Obama administration released the Reauthorization of the ESEA that further supported the reforms of the ARRA. This blueprint for reform specifically required that school districts implement an evaluation system that differentiated teachers according to their effectiveness based on multiple measures that included student growth and provided meaningful feedback for improvement (U.S. Department of Education, 2010).

At the state level, educational reform efforts in Florida were also focused on teacher quality (Florida Government, 2011). On March 24, 2011, the Student Success Act, Senate Bill 736, was signed into law by Governor Rick Scott. This law tied merit

pay to student achievement. The intent of the law was to ensure that every classroom was taught by a highly effective teacher.

As a result of research reiterating the importance of effective teachers, teacher evaluation gained increased emphasis since 2009 (Sartain, Stoelinga, & Brown, 2011). According to Sartain et al. the majority of evaluation systems are not effective because they do not provide teachers with timely and valuable suggestions for improving instruction and are often based on a single principal observation. Sartain et al. went on to say that many evaluation instruments are meaningless due to their subjectivity and their failure to differentiate between strong and weak instruction. Darling-Hammond (1996) also noted that most evaluation systems do not successfully recognize poor teachers. According to Darling-Hammond, diverse school systems such as Chicago, San Francisco, Atlanta, and Denver seldom dismiss teachers who have poor performance. In fact, in a year's time less than one percent are dismissed. In response to weaknesses with the evaluation systems, policymakers are demanding that student learning growth be used to evaluate teachers (Sartain et al., 2011).

According to the Student Learning, Student Achievement Task Force convened by the National Board for Professional Teaching Standards (2011), the quality of a teacher should be linked to student learning. It is feasible for measurements of teacher quality to include systematic evidence of student learning. Because learning is a collective effort and many variables beyond the control of the teacher can affect it, an evaluation system that includes both the act of teaching and student learning provides a more accurate measurement of teacher effectiveness (Tucker & Stronge, 2005). The Student Learning, Student Achievement Task Force stated that "Gains in student learning

must always be examined within the context of teaching practice to ensure that they are connected to what teachers are doing in the classroom” (National Board for Professional Teaching Standards, 2011, p. 9).

In studying rigorous teacher evaluation systems, Baum (2011) concluded that classroom practices contributed to student learning. Baum further found that a teacher’s effectiveness not only increased during the evaluation process, but the effect sizes in the years following the evaluation were even larger. According to Baum, evidence to date suggests that a good evaluation system has the capacity to improve teacher effectiveness. Although the basis of all evaluation systems is to help teachers to improve, the components to accomplish this vary. The question regarding what evaluation components provide the best measurement of teacher effectiveness tied to student learning is still unclear.

Although federal and state policymakers provided most of the impetus for new teacher evaluation systems, the task of designing the instruments falls to the school districts (Glazerman, Goldhaber, Loeb, Raudenbush, Staiger, Whitehurst, & Croft, 2011). Glazerman et al. described the changes to the evaluation systems as follows:

A new generation of teacher evaluation systems seeks to make performance measurement and feedback more rigorous and useful. These systems incorporate multiple sources of information, including such metrics as systematic classroom observations, student and parent surveys, measures of professionalism and commitment to the school community, more differentiated principal ratings, and test score gains for students in each teacher’s classroom. (para. 2)

This study will focus on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument and determine the relationship that exists between a teacher’s ratings on the professional practices components and their value-added measurement (VAM) score.

## Instrument Development

A performance appraisal project team was developed in Brevard Public Schools for the 2010-2011 school year with the purpose of “writing a new system for performance appraisals for instructional and school-based administrators” that complied with the parameters as set forth by Race to the Top (RTTT) and Senate Bill 736 (J. Salamone (personal communication, April 2, 2012)). The project team was led by the Director of Human Resources & Labor Relations and the Central Area Superintendent. In addition to the team leaders, the members included four principals (two elementary, one middle school, and one high school), four teachers (two elementary, one middle school, and one high school), three directors (Career and Technical Education, Administrative Support Services, and Professional Development), one area superintendent, and one human resource analyst. Sixteen meetings were held throughout the year to share research findings relative to evaluation systems and design the system. Two of the meetings were visits to Ocala and Hillsborough, Florida, to learn about their evaluation systems. A share site was developed on which all meeting agendas, minutes, and team documents were housed to share with the public. Brevard County’s instructional practices rubrics were reviewed by Charlotte Danielson, the founder of the Framework for Teaching, and the entire appraisal system was presented “to about 1,000 teachers and Leadership Team members in area meetings and a selected focus group to gain feedback” (J. Salamone (personal communication, April 2, 2012)).

## Purpose of the Study

The purpose of this study was to determine the relationship between the professional practices components of the School Board of Brevard County Instructional

Personnel Performance Appraisal System Instrument and student achievement as measured by a teacher's value-added measurement (VAM) score. The desired outcome and broader purpose was to determine the construct validity of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument so that the professional practices components could be further developed to enhance student learning. This study was also intended to provide Brevard Public Schools with input for further revisions to the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument.

#### Statement of the Problem

Because teacher quality is so closely linked to student achievement, it is critical that school districts develop evaluation instruments that not only measure student achievement but also provide teachers with feedback that can be used to improve the quality of their instruction (Sanders & Rivers, 1996). With the 2011-2012 school year being the initial year of implementation for the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, limited information existed as to the quality of the instrument as determined by its construct validity. According to Barrette, Morton, & Tozcu (1995), the validity of an instrument is a function of its purpose. One purpose of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument was to “measure quality instructional performance essential for promoting high student achievement and increased instructional improvement” (Brevard Public Schools, 2011, p. 7). Being the evaluation instrument's first year of implementation, nothing was known about how the professional practices portion of the instrument related to student achievement.

Therefore, the overarching question was whether or not the evaluation instrument was serving the purpose that it was intended. For the purpose of this study, a teacher's value-added score was the measurement used to determine student achievement.

### Research Questions and Hypotheses

The problem statement can be summarized by the question "To what extent, if any, is there a relationship between the professional practices component ratings and the value-added assessment score on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument?" The study was guided entirely by the following research questions:

1. To what extent, if any, is there a relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

2. To what extent, if any, is there a relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

3. To what extent, if any, is there a relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard

County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

4. To what extent, if any, is there a relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

5. To what extent, if any, is there a relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

6. To what extent, if any, is there a relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

7. To what extent, if any, is there a relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

8. To what extent, if any, is there a relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School



## Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

### Delimitations

This study was delimited to the reporting of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument ratings for Brevard Public School instructional personnel assigned to Grades 4-10 who taught reading and/or mathematics measured by the Florida Comprehensive Assessment Test (FCAT) during the 2011-2012 school year and received teacher aggregated VAM scores. Instructional personnel who taught mathematics assessed with either the Algebra I or Geometry End of Course (EOC) exam were excluded from this study. This included the following courses: (a) Algebra I, (b) Algebra I Honors, (c) Algebra IB, (d) International Baccalaureate Middle Years Program Algebra, (e) Pre-AICE Mathematics 1, (f) Geometry, (g) Geometry Honors, (h) International Baccalaureate Middle Years Program Geometry, and (i) Pre-AICE Mathematics 2 (Florida Department of Education, 2012).

The study was further delimited by not including the instructional personnel at the charter schools, the two adult centers, the alternative learning centers, virtual schools, and the school board office. Furthermore, this study only addressed the evaluation components as outlined on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. The study also evaluated student learning based on those students enrolled with a teacher during the entire 2011-2012 school year and those students who had prior year FCAT data.

## Limitations

The following factors were limitations of the research:

1. the lack of reliability and validity measures associated with the instruments used to gather the data because of them being used for the first time,
2. the lack of perceptual consistency and or distortions among evaluators,
3. the lack of accuracy of data due to human error since many ratings were manually input,
4. the lack of generalization to other school districts that have different teacher performance appraisal instruments, and
5. the lack of three years of performance data in determining a teacher's aggregated value-added score.

## Definitions of Terms

For the purpose of this study, the following terms were utilized:

1. **Aggregated Value-Added Scores:** One score that represents the total value-added score of a teacher regardless of his grade level or subject area (Schafer, 2012).
2. **Charlotte Danielson's Framework for Teaching:** A set of teacher responsibilities including 22 components grouped into four domains that studies have identified as being related to student learning (Strong, 2011).
3. **Domain:** A category for the classification of teacher performance standards (Stronge & Tucker, 2003).
4. **Florida Comprehensive Assessment Test (FCAT):** A Florida statewide assessment administered in Grades 3-10 to students to determine their understanding of the

- Next Generation Sunshine State Standards (Florida Department of Education, 2011).
5. Formative: Judging performance for improvement purposes (Stronge & Tucker, 2003).
  6. Goals and Roles Evaluation Model: A hybrid evaluation model developed by James Stronge that utilizes a combination of approaches from multiple models (Stronge & Tucker, 2003).
  7. Individualized Growth Plan: “Teacher learning that is grounded in day-to-day teaching practice and is designed to enhance teachers’ content-specific instructional practices with the intent of improving student learning” (Croft, Coggshall, Dolan, Powers, & Killion, 2010, p. 2).
  8. Performance Indicators: Typical observable or documentable behaviors that exemplify the degree to which a teacher is fulfilling a performance standard (Stronge & Tucker, 2003).
  9. Professional Learning Community: “A community with the capacity to promote and sustain the learning of all professionals in the school community with the collective purpose of enhancing student learning” (Bolam, McMahon, Stoll, Thomas, & Wallace, 2005, p. 145).
  10. School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument: The instrument used by Brevard County administrators to rate the performance of instructional personnel (Brevard Public Schools, 2011).
  11. Summative: Judging performance for accountability purposes (Stronge & Tucker, 2003).

12. Teacher Effectiveness: A teacher's ability to sustain or accelerate the academic growth of a student as measured by value-added measures (Goe, 2007).
13. Teacher Performance Standards: A teacher's professional responsibilities (Stronge & Tucker, 2003).
14. Teacher Quality: Teacher qualifications; characteristics such as attitudes, beliefs, and attributes; classroom practices; and student learning growth (Stronge & Tucker, 2003).
15. Value-Added Measurement: Using students' current and historical test scores to estimate the effect that a teacher has on their learning growth while controlling for various covariates (Hill, Kapitula, & Umland, 2010).

### Theoretical Framework

Evaluation models vary across the country in local school systems; however, many school districts utilize a combination of approaches taken from multiple models (Stronge & Tucker, 2003). James Stronge developed such a hybrid evaluation model called the Goals and Roles Evaluation Model (Stronge & Tucker, 2003). According to Stronge and Tucker (2003), this theoretical model was based on planning and assessment at the organizational level, teacher role expectations, feedback on performance, and improvement. What distinguished this evaluation model from most others was that it focused not only on individual improvement, but also on its link to organizational improvement as a collaborative effort through the collective performance of all professional stakeholders (Stronge & Tucker, 2003). As such, "A quality evaluation system should encourage the improvement of professional educators as a means of improving the school system as a whole" (Stronge & Tucker, 2003, p. 24).

Stronge and Tucker (2003) contended that the purpose of this model was both formative and summative assessment. This model was formative in that it focused on the improvement of teachers through constructive feedback, and summative in that it focused on the performance of teachers to assure that they were being accountable for their roles. Stronge and Tucker also pointed out that this model was specifically designed to contribute to the organizational goals, improve instruction through performance accountability and professional growth, and establish a collaborative process between the evaluator and the teacher. The Goals and Roles Evaluation Model was characterized by a comprehensive language, adaptability to various educational roles, standardization across the entire system, emphasis on communication during the total process, and utilization of multiple data sources (Stronge & Tucker, 2003). Figure 1 shows the framework of the Goals and Roles Evaluation Model. Permission to reproduce this figure was granted by Dr. James H. Stronge on March 21, 2012 (see Appendix A for Permission to Reproduce).

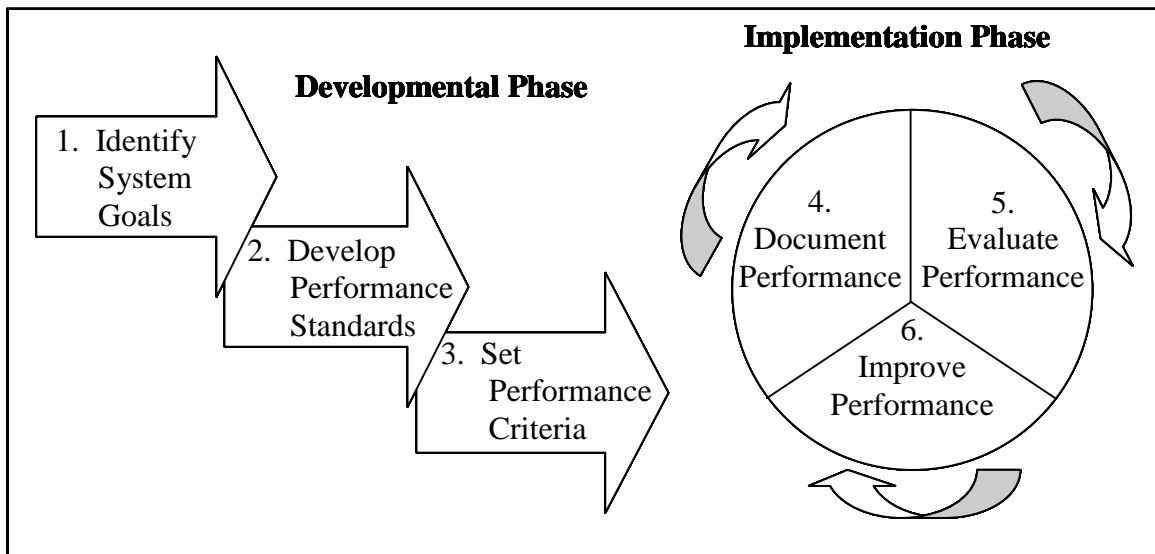


Figure 1: Goals and Roles Evaluation Model (Stronge & Tucker, 2003, p. 30).

According to Stronge and Tucker (2003), the Goals and Roles Evaluation Model was comprised of six steps. The first step was to identify the system goals. Stronge and

Tucker believed that a prerequisite to developing an evaluation system was to identify organizational goals to assure the alignment of the goals on which the performance evaluation system was based.

The second step of the Goals and Roles Evaluation Model as identified by Stronge and Tucker (2003) was to develop performance standards. Stronge and Tucker insisted that job performance standards were based on expectations of what teachers should be able to do. Furthermore, Stronge and Tucker claimed that the standards must be clear and easily understood by everyone involved in the process as well as able to be observed and measured. Accordingly, Stronge and Tucker believed that performance standards were the foundation on which the evaluation system was built. The performance responsibilities were detailed through a three-tier description including domains, performance standards, and performance indicators (Stronge & Tucker, 2003). Domains were categories for the classification of teacher performance standards. Performance standards were the duties and job responsibilities of the teacher. Performance indicators were typical observable or documentable behaviors that exemplified the degree to which a teacher was fulfilling a performance standard. The state's Race to the Top plan requires that the Florida Educator Accomplished Practices, established in 1998 by the State Board of Education and revised on December 17, 2010, be the basis for teacher performance in teacher appraisal systems (Haithcock, 2011).

The Goals and Roles Evaluation Model Step 3 as identified by Stronge and Tucker (2003) was to set performance criteria. According to Stronge and Tucker (2003), "Evaluation is a process, not an event" (p. 44). Therefore, there should be ongoing discussion between the evaluator and the teacher so that strengths can be reinforced and

weaknesses improved. Stronge and Tucker recommended using a three- or four-point rating scale because it offered enough differentiation so professional growth was encouraged yet it was not so intense that it was difficult to distinguish among the ratings. The use of a performance appraisal rubric more clearly defined the performance ratings in practical terms and assisted evaluators with determining how well a teacher performed on each performance standard (Stronge & Tucker, 2003).

Stronge and Tucker (2003) identified the fourth step of the Goals and Roles Evaluation Model as documenting performance. Stronge and Tucker believed that multiple data sources must be used to document performance. Basing an evaluation on only one data source, such as the common observation, was unreliable since it reflected only a small sample of actual teaching time, might be an artificial reflection of what was taking place on a daily basis, did not reflect all of the teacher responsibilities beyond instruction, and measured processes only and not results (Stronge & Tucker, 2003). Data sources upon which evaluations may be based include observations, teacher portfolios, client surveys, and student performance data. Stronge and Tucker found that documenting performance using multiple data sources increased validity, increased reliability, decreased subjectivity, increased the teacher and evaluator's comfort level, and created a more realistic picture of a teacher's performance.

The Goals and Roles Evaluation Model Step 5, according to Stronge and Tucker (2003), was to evaluate performance. Stronge and Tucker found that in most states certain parts of the process were mandated by the legislature and the specifics were left up to local school boards. Stronge and Tucker strongly believed that who will conduct the evaluations, how often and when the evaluations will occur, and how the evaluations

will be done were three questions that needed to be clearly answered upfront. Stronge and Tucker recommended that more than one person judge the performance of a teacher, and evaluation timelines be aligned with state requirements, negotiation agreements, and school board policy.

The final step in the Goals and Roles Evaluation Model as outlined by Stronge and Tucker (2003) was improving performance. According to Stronge and Tucker, this step was not intended to be a stopping point but rather a springboard into the next cycle. As stated by Stronge and Tucker, in order for an evaluation to be meaningful and not just an isolated event, it must be related to staff development and overall school improvement. The most prevalent strategy for facilitating performance improvement was conferencing including feedback, reflection, goal setting, and improvement assistance (Stronge & Tucker, 2003).

### Summary

Realizing that the most important factor in student achievement is the quality of the teacher, it is critical that school districts have an effective method in place not only for identifying teacher effectiveness but also for improving the quality of teachers. Designing an evaluation system that serves the function of both accountability and improvement is a difficult task. With student improvement of learning being the ultimate goal of any evaluation system, it is vital that there be a strong relationship between the various evaluation components and student learning. Adding to the research on this topic can only bring educators one step closer to identifying the right measures to include in an evaluation system with the purpose of increasing teacher effectiveness.



## Organization of the Study

Chapter 1 will provide an introduction including the purpose of the study along with its statement of the problem and the resulting research questions and their hypotheses. Delimitations and limitations as well as definitions of terms will be presented. This chapter will conclude with a theoretical framework on which the study is based and an overview of the methodology. Chapter 2 will present a review of literature including research associated with the problem. Chapter 3 will include the methodology used to perform the study including an overview of the research questions and hypotheses; the research design, describing the population, data collection, and procedures for data analysis; and a summary. Chapter 4 will provide the results of the statistical tests on the data. Chapter 5 will present an overview of the findings, a statistical analysis of the results, and future recommendations as a result of the findings.

## CHAPTER II LITERATURE REVIEW

This chapter provides the justification necessary for conducting a study on the relationship between the professional practices component ratings and the value-added assessment rating on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. More than 400 empirical studies have been conducted to determine what, if any, teacher qualifications predict student achievement (Schacter & Thum, 2004). The results were best summed up by Hanushek (2002) when he said that a teacher's qualifications have no effect on their quality, and in order to improve the quality of teachers policymakers must focus on teacher performance. "By identifying, describing, and categorizing the dispositions, knowledge, and instructional skills of effective teachers, one can develop interventions to improve teaching and to reduce the variability in student achievement gains between classrooms" (Schacter & Thum, 2004, p. 412). Research has shown that the greatest probability of improving student learning is to implement an accountability system in which teacher quality and productivity are integrated (Schacter & Thum, 2004).

In order to review all available research relevant to the study, the following six components were identified: (a) teacher quality, (b) historical perspective (c) observation, (d) individualized growth plans, (e) collaboration and mutual accountability, and (f) value-added assessment. The terms relative to each component were defined along with related synonyms. These terms were narrowed or broadened based on the

results of the search. Prevalent researchers associated with each component were used to further search the literature. A detailed overview of the search parameters is provided in Appendix B. Except for the historical perspective literature review search, limitations to the search were set to include studies within the last ten years carried out in educational public school settings located within the United States. Electronic databases used to search the literature were comprised of ERIC, ProQuest, Dissertations and Theses, and Internet sources. Multiple sources were referenced within each database including journals, books, working papers, policy briefs, research reports, executive summaries, PowerPoints, and webcasts. Once the research was identified, the abstracts were examined or the content scanned to select those investigations most closely related to the study.

This review of literature provides a framework for the study by defining teacher quality, a discussion of the history of the evaluation system in the United States, and a discussion of the evaluation components relative to this study and the research supporting their validity. Each section concludes with a summary of the research that attempts to determine the correlation that exists among observations, individual growth plans, mutual collaboration and accountability, and the value-added scores that teachers receive.

### Teacher Quality

To design a valid teacher evaluation instrument to measure the quality of a teacher, teacher quality must first be clearly defined (Goe, Bell, & Little, 2008). According to Goe (2007) the definition of teacher effectiveness and how it is measured has changed substantially over the past 30 years. There is still disagreement with regards to its definition. Strong (2011) found that the description of teacher quality varied

throughout the literature depending on a person's viewpoint. Strong's literature review indicated that definitions of teacher quality focused on competencies; such as, certification and experience, personal qualities; such as, compassion and fairness, standard of pedagogy; such as classroom management and instructional strategies, and student learning outcomes. Hinchey (2010) distinguished among teacher quality, teacher performance, and teacher effectiveness. According to Hinchey, teacher quality is defined as characteristics of a teacher including experience, education and beliefs; teacher performance is a teacher's behavior inside and outside of the classroom; and teacher effectiveness is a teacher's impact on student learning. Munoz and Chang (2007) found that the characteristics of a teacher in the classroom are interrelated to student growth. Heneman, Milanowski, Kimball, and Odden (2006) found that there is a positive relationship between students' standardized performance scores and student learning as influenced by instructional practices. Hinchey (2010) recommended that teacher effectiveness be combined with teacher performance and teacher quality measures when assessing a teacher.

Goe (2007) defined teacher quality in terms of inputs, processes, and outcomes. Inputs included teacher qualifications and characteristics such as beliefs, attitudes, and attributes. Processes included teacher classroom practices, and outcomes included teacher effectiveness determined by student learning growth. Goe cautioned using any one of the three strands in isolation to define teacher quality because of their interrelatedness. See Figure 2 for a graphic representation of Goe's teacher quality framework. Permission to reproduce this figure was granted by Dr. Laura Goe on March 11, 2012 (see Appendix C for Permission to Reproduce).

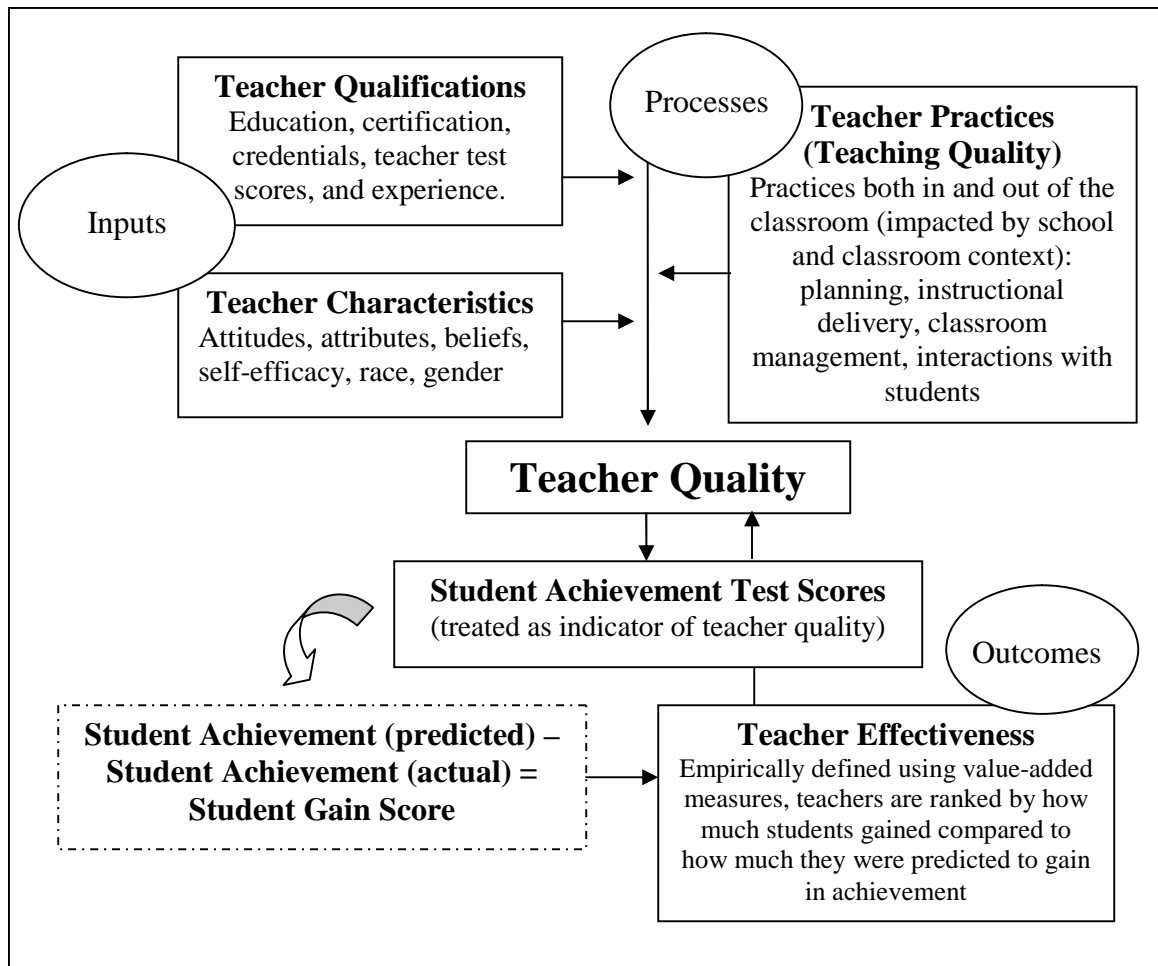


Figure 2: Teacher Quality Framework (Goe, 2007, p. 9).

Goe et al. (2008) also evaluated the research on teacher effectiveness and sought the advice of teacher quality experts to arrive at the following comprehensive definition that goes beyond student achievement gains:

- Effective teachers have high expectations for all students and help students learn, as measured by value-added or other test-based measures, or by alternative measures.
- Effective teachers contribute to positive academic, attitudinal, or social outcomes for students such as regular attendance, on-time promotion to the next grade, on-time graduation, self-efficacy, and cooperative behavior.
- Effective teachers use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapting instruction as needed; and evaluate learning using multiple sources of evidence.
- Effective teachers contribute to the development of classrooms and schools that value diversity and civic-mindedness.

- Effective teachers collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly the success of students with special needs and those at high risk for failure. (p. 8)

Although there exists a variety of views on the components that define an effective teacher as well as what they are called, it was evident from the literature that multiple components do exist. A report prepared by The New Teacher Project (2010) indicated that a teacher's performance cannot be accurately portrayed by a single piece of data; and therefore, multiple measures must be included as part of any evaluation system in order to effectively judge a teacher's performance. The literature supports teacher evaluations based on the use of multiple data sources (Peterson, 2004).

#### Historical Perspective

In order to fully understand the dysfunctional nature of the teacher evaluation system, it is necessary to review its history (Hazi & Rucinski, 2009). According to Hazi and Rucinski, the personnel function associated with teacher evaluations began in the early 18<sup>th</sup> century. Committees of school personnel, selectmen, ministers, and later administrators and superintendents were tasked first with the inspection of equipment, facilities, and student achievement and then with teachers' methods of instruction (Hazi & Rucinski, 2009). According to Kovats (2006), teacher evaluations date back to the 1800s during the time period of one-room school houses. Common people supervised the schools taught by qualified teachers based on their ability to read, write, and handle strong farm boys who might be intransigent (Borthwick, Cohodes, Sennette, & Touhey, 2009). The purpose of teacher evaluations was related to promotion and salary increases with localized standards aligned with educational objectives (Kovats, 2006). Borthwick et al. (2009) reported that in the late 1800s predominantly in urban areas, schools were

contracted by local educational governments. Teachers were hired based on their political affiliation and not their competence; therefore, resulting in the absence of evaluation methods.

Observations done through one-time classroom visits took place from 1910 through the 1930s (Hazi & Rucinski, 2009). This practice consisted of quietly taking notes in the back of the classroom as the teacher taught followed by a conference focusing first on commendations and then on criticisms. The early 1900s were also accompanied by the development of Fredrick Taylor's scientific management theory and an attempt at quantifying learning (Connor, 1920). Influenced by the efficiency movement of Frederick Taylor, rating scales to measure teacher effectiveness were introduced based on the assumption that if descriptors could be developed by scientists for effective teachers, then ineffective teachers could be identified and targeted for improvement (Hazi & Rucinski, 2009). Teacher rating scales based on multiple factors gained widespread popularity in 1915 when a study was published in a yearbook of the National Society for the Study of Education that proposed the use of a 45-item rating scale to rate teachers that included both effective teaching and personality traits (Medley & Coker, 1987).

Superintendent William Connor of the Republic, Michigan schools in 1917 was the first to use a rating scale to judge teacher efficiency and connect it to additional teacher compensation (Shaffer, 1990). According to Connor (1920):

It is the teacher's task to make changes for the better in the abilities, habits, and attitudes of boys and girls. Her efficiency can be evaluated fairly only in terms of her success at this task. In other words, if a teacher is rated at all, she should be rated, not by the clothes she wears, or the method she chooses, but by the results she secures. (p. 338)

Although the results indicated that in a three-year period teacher quality more than doubled, the program ended in 1919 because of the lack of funding secured from the public (Connor, 1920).

The first study correlating teacher ratings to student achievement was published in 1921 (Hill, 1921). The superintendent of Marengo, Illinois schools, C. W. Hill, conducted a study across the United States surveying school systems to determine what type of evaluation rating systems were being used. Results indicated that 41% of the schools surveyed used some type of rating scheme to determine teacher efficiency. Based on established criteria, Hill selected 135 teachers from the survey respondents in the following school systems to take part in his study: Winnetka, Illinois; Gary, Indiana; and Detroit, Michigan. Hill correlated the teachers' efficiency grades with standardized test scores for arithmetic, penmanship, and spelling. Hill concluded that the correlation between test scores and ratings in all fields in all schools averaged .454. According to Medley and Coker (1987), from 1930 to 1955, 11 additional studies conducted to determine if a correlation existed between the rating scales of teacher performance and student achievement concluded that the correlation was near zero.

Prior to the 1950s, Daley and Kim (2010) recall that personal traits were judged according to moral and ethical standards to determine the quality of a teacher. Danielson and McGreal (2000) also found that teacher evaluations focused on presage variables, or personal attributes, because a belief existed among educators that teacher effectiveness was based on the existence of these traits.

Teacher unions were launched in the 1950s resulting in collective bargaining agreements that delineated teacher evaluation processes (Borthwick et al., 2009).



According to Kovats (2006), unions began to influence the evaluation process by establishing teacher criteria and dismissal and promotion guidelines although local school boards still assumed control over the process. In the 1960s evaluations shifted to focus on the behavior of teachers in the classroom, and by the end of the 1960s, with the aim at increasing the evaluation instrument's objectivity, observations were based on pre-identified categories (Barrette et al., 1995). Although the Equality of Educational Opportunity, also known as the Coleman Report, published in 1966 tried to minimize the effect of the teacher on student outcomes and emphasize the impact of ethnicity, socioeconomic status, and family background as the major variables influencing student achievement, many researchers continued to dispute these findings (Shaffer, 1990).

The primary purpose of evaluations during the time period prior to the 1970s was summative aimed at employment decisions with minimal focus on the improvement of teaching (Danielson & McGreal, 2000). During the 1970s teacher evaluations gained in popularity as a result of public pressure for accountability, and an influx of research in the field of teaching contributed to a shift to observations and rating instruments to measure what was taking place in classrooms (Liu, 2010). It was not until the 1980s and 1990s that school reform focused on using teacher evaluations to improve instructional quality (Daley & Kim, 2010). Daley and Kim further noted that it was also during this time period that many states were passing laws aimed at systematizing the implementation of teacher evaluations.

In the 1980s accountability of the schools and teachers was paramount. The Nation at Risk report was published in 1983 by the National Commission on Excellence in Education, and it emphasized how the education system in the United States was

failing and producing graduating students who were not prepared to compete nationally in the workforce (National Commission on Excellence in Education, 1983). In response to a Nation at Risk, evaluations were mandated by state legislatures and were the primary measure of accountability (Hughes, 2006). According to Daley and Kim (2010), during this time period, the number of states enacting laws regulating the execution of teacher evaluations was increasing. Teacher evaluation reform connected to merit pay flourished to address mediocrity that existed among the nation's schools and teaching workforce (Donaldson, 2009). Despite this initial reformation, Donaldson found that when the economy resulted in teacher lay-offs, it was seniority not evaluation results that was the basis of decisions. According to Donaldson, when the dollars necessary to support evaluation reform decreased, teacher evaluation experimentations subsided and curriculum standards become the new priority.

A 1996 publication by the National Commission on Teaching and America's Future entitled "What Matters Most: Teaching for America's Future," brought teacher quality center stage of most political agendas (Hughes, 2006). In 2001, with the passage of No Child Left Behind, states continued to strengthen their control and regulation of evaluation practices at the local level by delineating teacher quality, standards for the training of evaluators, and data collection requirements (Daley & Kim, 2010). Daley and Kim further stated that the resulting evaluation systems were not designed to assist teachers in growing professionally but were infrequently performed to meet bureaucratic requirements. Weisberg, Sexton, Mulhern, and Keeling (2009) also agreed with this statement. They surveyed 15,176 teachers in 12 school districts and found that 75% of the teachers had received no feedback specifically on how their instructional practice

could be improved. Although classroom visits and post conferences continued to expand over the years, according to Black (1993), they still for both the teacher and principal remained an unpleasant, time consuming task.

In response to No Child Left Behind and its focus on highly qualified teachers in every school, the focus on teacher evaluation as a policy target brought about many changes (Hazi & Rucinski, 2009). The National Governors Association (NGA) targeted teacher evaluation as a tool for the improvement of instruction. In a NGA policy brief, Goldrick (2002) summarized anticipated changes in evaluation systems:

A purposeful evaluation system measures teaching outcomes, not simply teaching behavior. Evaluations that are well-designed and integrated with curriculum and professional standards can accomplish more than assuring basic competence. They can help states and districts measure the effectiveness of teachers at various points in their careers, identify highly skilled teachers, offer specific recommendations to improve teaching, inform professional development, and demonstrate accountability for student achievement. State policymakers should treat teacher evaluation as an integrated component of a comprehensive strategy to improve overall teaching quality. (p. 2)

Over the last decade, teacher quality continued to be the focus of evaluation systems (Weisberg et al., 2009). The New Teacher Project (TNTP), a national nonprofit organization, addressed the reality of poor evaluations nationwide in stating that teacher effectiveness, a key to student achievement, was not used as a basis for meaningful decisions in schools (Weisberg et al., 2009). In their report, Weisberg et al. described the evaluation process as a Widget Effect where all teachers' classroom effectiveness was judged the same with no distinctions being made between the good and the poor teachers.

In response to the need to focus on teacher quality through effective evaluation systems, the American Recovery and Reinvestment Act (ARRA) was signed into law by President Obama on February 17, 2009 (U.S. Department of Education, 2009). This act

made \$4.35 billion available for a competitive grant program referred to as “Race to the Top.” Points were awarded based on various criteria with the highest single points being awarded for improving principal and teacher effectiveness. This reform effort was an attempt to assist states in developing evaluation systems that linked teacher and student data.

The Reauthorization of the ESEA was released on March, 13, 2010, by the Obama administration to further support the reforms of the ARRA. This reform blueprint required that school districts specifically implement an evaluation system in which teachers were differentiated not only on multiple measures that included student growth but also were provided meaningful feedback for improvement (U.S. Department of Education, 2010). To date, states and school districts across the country are restoring their evaluation systems by passing supportive legislation and collective bargaining agreements (Daley & Kim, 2010).

### Observation

Darling-Hammond, Wise, and Pease (1983) described the methods used for teacher evaluations as typically not reliable or valid. Likewise, Peterson (2000) found in his review of evaluation literature that the average practices associated with teacher evaluation do not improve teachers or provide an accurate representation of what is happening on a regular basis in the classroom. Danielson and McGreal (2000) criticized evaluation systems as shallow with little relevance for the enhancement of student learning. According to Milanowski (2004), “As a measurement process, the reputation of teacher evaluation is not particularly good” (p. 34). Because Donaldson (2009) found that 60% of the school districts in the United States use some form of observation in

teacher evaluations, it is important for this study to determine what the research says concerning the relationship between observation scores and student learning.

The goal of classroom observations is to improve students' academic outcomes by assisting teachers in perfecting their practice (Kane & Staiger, 2012). Kane and Staiger further stated that classroom observation instruments that do not correlate to student outcomes do not contribute to the achievement of this goal. An observation instrument that is valid must be closely aligned with student achievement gains according to Kane and Staiger. "Testing for validity means determining the extent to which teaching indicators are related to student outcomes" (Kane & Staiger, 2012, p. 22).

Hundreds of studies have been conducted since the 1970s using classroom observations (Waxman, 2011). According to Waxman, the main purpose of teacher observations is to improve instruction. Feedback assists teachers in identifying their strengths and weaknesses ultimately leading to instructional improvement. It was not; however, until 1992 that the first value-added model was created by Dr. William L. Sanders and implemented in the Tennessee Public Schools (Shurtleff & Lored, 2008). Shurtleff and Lored further pointed out that as of 2007, value-added assessment was officially mandated in only four states yet 15 states utilized value-added assessment in some form. Therefore, studies using value-added measures as indicators for student achievement were limited.

#### *Los Angeles Elementary Charter School*

Gallagher (2004) studied the validity of a teacher evaluation system at a Los Angeles elementary charter school servicing 1,200 students all of who were on free or reduced lunch. The evaluation system was performance based with subject-specific

evaluations written by the staff based on Charlotte Danielson's Framework for Teaching. Danielson's Framework for Teaching is a set of teacher responsibilities including 33 components grouped into four domains that studies have identified as being related to student learning (Strong, 2011). In this study, teachers were evaluated on up to ten domains, seven of which were related to a specific subject. Each domain had up to nine standards. The system was used for both formative and summative assessment purposes. Teachers were evaluated three times a year by peer evaluators and administrators who had been extensively trained. Inter-rater reliability was high with an alpha coefficient of .86. The peer evaluator and administrator's scores along with a self-evaluation score were evenly weighted in determining a composite score. The sample for this research included 34 second- through fifth-grade teachers with a student sample size of 584 for reading and mathematics, 532 for language arts, and 527 for the composite. The correlation between teacher evaluation scores and student achievement were statistically significant in reading with a coefficient of .50 at a .01 alpha level and in the composite score with a coefficient of .36 at a .05 alpha level. Teachers' evaluation scores in reading were determined to be high predictors of student performance. For example, an increase of one point in teacher evaluation scores in literacy resulted in a 14 point increase in student performance. Teacher evaluation scores accounted for 34% of the variation in reading but only 3% in mathematics, 1% in language arts, and 13% in the composite score. According to Gallagher (2004), student achievement scores varied significantly based on classroom effects. Qualitative results indicated that the limited knowledge of mathematics instruction resulted in the inability of both teachers and evaluators to distinguish among the varying degrees of teacher quality in mathematics.

*Washoe County, Nevada School District*

Borman and Kimball (2004) conducted a study in Nevada Washoe County School District where teachers were evaluated annually by an administrator using an observation instrument modeled closely after Charlotte Danielson's Framework for Teaching. This study targeted 4,676 students and 266 teachers in Grades 4-6. The population had a minority rate of 34% and a free and reduced lunch rate of 25%. The composite scores for each of the four domains included in the observation were averaged together to arrive at a teacher quality rating. Student achievement results were based on district criterion referenced assessments in Grades 4 and 6, state criterion referenced tests in Grades 4-6, and the Comprehensive Test of Basic Skills, Terra Nova, in Grade 4. Using a value-added model to determine teacher effect, this study concluded that the average classroom achievement for a teacher whose teacher quality rating was at the 84<sup>th</sup> percentile of the distribution of scores was .2 standard deviations higher than the average classroom mean in the district.

*Arizona*

A study conducted by Schacter and Thum (2004) assessed teaching quality using an observation instrument comprised of 12 teaching performance standards with corresponding descriptive narratives based on research related to behaviors, models, strategies, and qualifications of teachers. The purpose of the study was to determine if the ratings associated with the standards predicted student achievement gains as measured by a value-added assessment model. The sample included 52 Grade 3-6 elementary school teachers in Arizona and 910 students with a 53% minority rate. All teachers were observed eight times over a nine-month period by trained graduate student

researchers and rated on a 5-point scale from exemplary to ineffective. Inter-rater reliability was high among raters ranging from .74 to .92. Outcome measures were based on the Stanford 9 Achievement Tests in reading, mathematics, and language. The results indicated that teacher quality and productivity were positively correlated at a level of 0.55 to 0.70. There was a difference of 4.75 standard deviations between the highest and lowest teacher quality ratings resulting in a difference in scale score of 32.3 in language, 33.5 in mathematics, and 22.5 points in reading. There was a difference of 1.61 standard deviations between the first and third teacher quality quartiles resulting in a difference in scale score of 10.9 in language, 11.4 in mathematics, and 7.6 in reading.

#### *Midsized Western United States School District*

Jacob and Lefgren (2008) conducted a study of 202 Grade 2-6 teachers in a midsized western school district in the United States to determine to what extent principals could identify effective and non-effective teachers. In terms of socioeconomic status and ethnicity, the school district was heterogeneous with a minority rate of 27% and a free or reduced lunch rate of 48%. The school district had an achievement level on the Stanford Achievement Test at the national mean. Core criterion-reference exams were used as value-added measures. Principals were asked to rate their teachers on a 10-point scale from inadequate to exceptional on 11 dimensions including overall effectiveness, work ethic, organization, classroom management, mathematics and reading achievement, role model, student and parent satisfaction, and relationship with colleagues and administration. The average of the ratings was 8.07 with 6.7 being the lowest principals' average. The ratings of principals within schools were normally distributed.



The results of Jacob and Lefgren's (2008) study indicated that the correlation between a teacher's value-added measure and the principal's rating of their ability to raise student achievement in reading and mathematics was .29 and .32 respectively. This study found that principals were able to correctly identify teachers in both the top and bottom quartiles based on their value-added measures in mathematics 70% of the time and in reading 55% of the time compared to the probabilities of 14% and 26% if ratings had been randomly assigned. Principals were not as successful in identifying teachers in the middle achievement distribution with probabilities of 62% as compared to 33% had the ratings been randomly assigned. According to Jacob and Lefgren, "Our findings provide compelling evidence that good teaching is, at least to some extent, observable by those close to the education process even though it may not be easily captured in those variables commonly available to the econometrician" (p. 130).

#### *Cincinnati Public School System*

According to Kane, Taylor, Tyler, and Wooten (2010), observation systems that are of quality are based on standards that are clear and objective, conducted by several evaluators who have been trained, and include multiple observations over time. The Cincinnati's Teacher Evaluation System (TES) includes peer evaluators who were trained to observe teachers three times throughout the year in addition to the one trained administrator's observation. The first observation was required to be announced whereas the subsequent ones were not. Teachers were observed annually until they obtained tenure status at which point they were observed every five years. The observation instrument used was based on Charlotte Danielson's Framework for Teaching including four domains, sixteen standards, and 32 elements. Each element was judged based on the

teacher's performance on a 4-point rubric scale of distinguished, proficient, basic, and unsatisfactory.

Milanowski (2004) examined the evaluation system in Cincinnati, Ohio, to determine if a relationship existed between the scores on the teachers' standards-based evaluations and their students' value-added measures. The sample was comprised of 212 teachers in Grades 3 through 8. A composite evaluation score was used for this analysis indicating the overall performance of a teacher and was calculated by adding the four domain scores. When the scores were correlated in reading, mathematics, and science across Grades 3-8, the combined coefficients were positive ranging from .43 in mathematics, to .32 in reading, to .27 in science. These results were considered a moderate correlation. According to Milanowski, the results suggested that beyond chance the Cincinnati evaluation system was successful at identifying teachers with students whose achievement levels were higher than expected. Milanowski went on to conclude that "Teacher evaluation scores may be useful as representations of teaching practices that affect student learning" (p. 49).

Six years later, a nine year study was conducted by Kane et al. (2010) in Cincinnati, Ohio, using longitudinal student data to analyze teacher observations and determine their relationship to student achievement. This study focused on only two of the four standards since they were the basis for the observation scores because the other two standards were evidence based. Using value-added estimates based on student achievement data from prior years, the teachers were separated into quartiles to determine if there was a difference in observation ratings of teachers in the highest quartile as compared to those teachers in the bottom two quartiles. Results indicated that teachers in

the upper quartile received higher ratings than those in the bottom two quartiles. Kane et al. found that teachers associated with high student achievement taught differently than teachers associated with low student achievement. In determining whether observation ratings could be used to predict student achievement growth, the results indicated that an increase of one point on the average observation rating was associated with a student achievement gain of .17 standard deviations in mathematics and .20 standard deviations in reading. Teacher observation ratings also improved on subsequent observation ratings on an average by .45 standard deviations no matter what the length of time among observations. Kane et al. concluded the following from this study:

Multiple alternative measures of teacher effectiveness may be more predictive of future student achievement effects than any single measure. This is true when classroom observation scores are brought into a model that previously only had student achievement measures. It is also true, perhaps more so, when student achievement measures are added to a model that only had classroom observation data. A teacher's past student achievement gains are a good predictor of future achievement gains, but measuring classroom practice likely improves the prediction. Teachers or administrators considering their future prospects for success should be open to including both forms of measuring past effectiveness. (pp. 28-29)

#### *A Measure of Effective Teaching Project*

A Measures of Effective Teaching (MET) project involving 3,000 volunteer teachers across the United States was carried out under the direction of the Bill & Melinda Gates Foundation to guide practitioners in structuring evaluation systems that improve teaching and learning (Kane & Staiger, 2012). The validity of five observation instruments was established based on value-added assessment measures as determined by students' performance on the Balanced Assessment in Math (BAM) and the Stanford 9 Open-Ended (SAT9 OE) reading assessment. The five classroom observation

instruments included: (a) Framework for Teaching (FFT), (b) Classroom Assessment Scoring System (CLASS), (c) Protocol for Language Arts Teaching Observations (PLATO), (d) Mathematical Quality of Teaching (MQI), and (e) UTeach Teacher Observation Protocol (UTOP). Teacher observation scores were compared to value-added scores across classes for the same teachers. This was done so that observer bias based on student behavior could be eliminated. The results of this study indicated a statistically significant positive correlation between classroom observation results and value-added scores for all classroom observation instruments. The difference in learning gains for teachers in the top quartile based on their value-added scores when compared to the bottom quartile teachers ranged from 2.6 months on the FFT to .6 months on the PLATO. When observation scores were combined with other measures such as student feedback and value-added scores on state tests, the relationship strengthened. For example, when the FFT results were combined with student feedback the 2.6 month quartile range difference increased to 4.8 months. When the FFT results and the student feedback results were combined with value-added scores on state tests, the 4.8 month quartile range difference increased to 7.6 months. Kane and Staiger concluded that the difference between low- and high-performing teachers when measured by three indicators was greater significantly than when measured by classroom observations alone.

#### *Northeastern Florida Schools*

A review of the literature showed that not all studies confirmed a relationship between objective performance and subjective ratings. Using data from three consecutive years, Aunchman (2009) conducted a study to determine if there was a relationship between Stanford Achievement Test (SAT) total reading and total mathematics scores of

Florida students in Grades 2-5 and teacher scores on the Teacher Skills Assessment Checklist, an observational instrument in which teachers were rated on effectiveness components. The instrument included 22 dimensions grouped into five domains. All dimensions were rated on a 4-point scale indicating mastery, professionalism, apprenticeship, or ineffectiveness. The sample for this study was comprised of 41 teachers from three different Florida Northeastern schools. Teachers were categorized into three groups based on their effect scores, and then their total checklist scores were correlated to their effect value-added scores. Results indicated that there was not a significant correlation between Teacher Skills Assessment Checklist scores and value-added scores ( $r = .001$ ,  $p = .986$ ). Further investigation indicated that the two dimensions that were related the most to student achievement were classroom management and the monitoring of student progress.

#### *Texas Independent School District*

In 2010, Pate conducted a study in the Independent School District in Texas to determine if a relationship existed between teacher effectiveness, as measured by the Professional Development and Appraisal System (PDAS), and fourth and fifth grade student achievement in reading and mathematics, as measured by the Texas Assessment of Knowledge and Skills (TAKS). This study had a sample size of 55 teachers who were scored on a 4-point scale; exceeds expectations, proficient, below expectations, and unsatisfactory, on 52 evaluation criteria in eight domains. Because Domain VIII focused on student performance, this domain was the portion of the observation instrument used for this analysis. The results of the study indicated that a relationship did not exist between the effectiveness of the teachers and student achievement as measured by the

PDAS and TAKS. The correlation was not statistically significant at the fourth grade level between teacher effectiveness and student mathematics performance ( $r = .09$ ,  $p = .68$ ) or student reading performance ( $r = .10$ ,  $p = .65$ ). Likewise, the correlation was not statistically significant at the fifth grade level between teacher effectiveness and student mathematics performance ( $r = .11$ ,  $p = .56$ ) or student reading performance ( $r = .15$ ,  $p = .47$ ).

### *Summary*

The review of literature indicated that a majority of the studies concluded that there is a positive relationship between observation rating scales and value-added scores. It must be noted that almost all of the studies found in the literature were conducted at the elementary level using a variety of assessments to determine VAM scores. Kane and Staiger (2012) emphasized that the success of classroom observations in improving student learning is contingent on quality implementation. Peterson (2004) said that the mere existence of behaviors as identified on a checklist or rubric do not equate with student learning; but rather the appropriateness of how these behaviors were performed.

### Individualized Growth Plans

The research on professional development was extensive and tied mostly to its inadequacies. Reform efforts from experts offered a dichotomy of solutions. Some researchers recommended that professional development be teacher specific and focused on what was happening in the classroom while others thought that professional development should be systemic (McLaughlin, 1990; Tye & Tye, 1984). Some researchers thought that the change associated with professional development should be

gradual while others thought that professional development based on a broader scope of change would result in better implementation (Fullan, 1985; McLaughlin & Marsh, 1978). According to Guskey (1997) reform efforts struggle in tailoring a sound professional development system as a result of these opposing views.

The literature on current professional development practices supported what Fullan (2001) found; lots of time had been wasted by teachers attending conferences and workshops that had resulted in unchanged classroom practices. In order to sustain the implementation of new classroom practices and meet the guidelines for a successful professional development program as outlined by Guskey and the National Center for Education Evaluation and Regional Assistance, individualized growth plans are replacing traditional staff development at the school level (Burke, 2000).

#### *Definition*

Individualized growth plans are a format of what the literature refers to as job-embedded professional development. Croft et al. (2010) defined it as “Teacher learning that is grounded in day-to-day teaching practice and is designed to enhance teachers’ content-specific instructional practices with the intent of improving student learning” (p. 2). Croft et al. further defined job-embedded professional development as a continuous improvement cycle where solutions are found by teachers for authentic instructional weaknesses. Job-embedded professional development is aligned with student achievement through the state standards and school improvement plan goals (Croft et al., 2010).

### *Underlying Assumptions*

Individualized growth plans are based on Sparks and Loucks-Horsley's (1989) model of individual staff development that stated:

Individuals can best judge their own learning needs and that they are capable of self-direction and self-initiated learning. It also assumes that adults learn most efficiently when they initiate and plan their learning activities rather than spending their time in activities that are less relevant than those they would design....The model also holds that individuals will be most motivated when they select their own learning goals based on their personal assessment of their needs. (p. 42)

In its review of professional development, the National Center for Education Evaluation and Regional Assistance concluded that key factors associated with successful professional development included it being specific to individual teachers' needs, based on student data, and built around a system in which time was allotted for practice and reflection (NGA Center for Best Practices, 2009). Guskey's (1997) meta-analysis of research on professional development led him to conclude that there were four principles upon which professional development programs tied to student achievement were based: (a) they had student learning as their goal, (b) they focused on organizational change in addition to individual change, (c) they required teachers to make changes that were small yet aligned with a larger vision, and (d) they had ongoing professional development embedded within the daily professional responsibilities.

According to Burke (2000) there are six distinct differences between the in-service model and the individualized growth plan model. The individualized growth plans are (a) ongoing, (b) determined by the teacher, (c) focused on professional growth, (d) focused on collaboration, (e) contextual, and (f) based on ownership. Burke further noted that the teacher determines the area of concern or new instructional strategy for the



professional development based on student needs and personal teaching goals aimed at improving the quality of teaching, expanding their knowledge base and skill repertoire, and gaining a better understanding of teaching and learning best practices. Practitioner research needs to be intentional and built upon a systematic plan including data that have been gathered, documented and reported (Barnatt, 2008).

Cochran-Smith and Lytle (1999) coined the term “inquiry as stance” to describe the process that teachers must use in order to successfully acquire knowledge to guide their practice. As Cochran-Smith and Lytle explained, the process is not an isolated occurrence but rather cyclical and recursive throughout a teacher’s career. Practitioner research is also collaborative in nature (Cochran-Smith, Barnatt, Friedman, & Pine, 2009). The questions on which inquiry of student learning are based must be broad (Barnatt, 2008). The result of the inquiry should be aimed at changing practices, providing new understandings, and prompting further inquiry (Barnatt, 2008).

According to Darling-Hammond and Bransford (2005):

Teachers learn to teach in a community that enables them to develop a vision for their practice; a set of understandings about teaching, learning, and children; dispositions about how to use this knowledge; practices that allow them to act on their intentions and beliefs; and tools that support their efforts. (p. 385)

Teachers must have a curricular vision, or a sense of direction, in their teaching (Darling-Hammond & Bransford, 2005). Darling and Bransford noted that this vision was critical along with the knowledge of best practices not only to help teachers reflect on their teaching but also to guide their professional practices.

Cochran-Smith et al. (2009) conducted a mixed method study of teacher candidates at Boston College to explore their practitioner inquiries. Forty-six inquiries were randomly selected and scored on a 100-point rubric. The results of the study

indicated that inquiries with high ratings were associated with teachers who possessed curricular vision as evident from their linkage of classroom interventions and instructional methods with their broader understandings of pupils as learners and classrooms as learning cultures. Cochran-Smith et al. (2009) further found that higher rated inquiries utilized multiple data sources to measure student learning, teaching strategies that were designed to address the full range of learning levels, and reflection that resulted in modifications to teaching practices. Finally, this study concluded that teachers who saw inquiry as recursive and understood that teaching was a continuous process of responding to data of practice had higher inquiry ratings (Cochran-Smith, et al., 2009).

### *Impact on Student Achievement*

#### *Quantitative Outcomes*

Dunaway, Do-Hong, and Szad (2010) concluded that “When systems invest in teacher knowledge and skills, a greater increase in student achievement occurs-more than with any other use of educational funds” (p. 6). Lawrence and the Florida State Department of Education (1974) in reviewing 97 in-service programs found that personalized programs with individualized objectives and activities were more successful in achieving their objectives than experiences that were the same for everyone. Geringer (2003) likewise stated that in school districts where individualized growth plans were being fully implemented coupled with reflection; an improvement in student performance was evident. Dunaway’s et al. (2010) review of literature on individual growth plans demonstrated that when individual growth plans were built around research-based

practices, the knowledge, skills, and performance of a teacher increased as well as student learning.

Although the National Center for Education Evaluation and Regional Assistance reviewed 1,300 studies in 2007 to determine if there was a correlation between professional development and student achievement, only nine of the studies could be used to draw conclusions that were reliable and valid (NGA Center for Best Practices, 2009). These nine studies concluded that professional development could impact student achievement and teaching practices. In fact the studies' findings indicated that 49 hours of professional development with a single focus could improve student achievement scores by as much as 21 percentile points.

Wallace (2009) studied the effects of professional development on instructional practices in third through eighth grade reading and mathematics in Tennessee and Connecticut to determine its effect on student learning gains. The results indicated that "The effects of professional development on teacher practice are small to moderate, with very small but occasionally significant indirect effects on student outcomes" (p. 591). Consistent in research findings was the fact that professional development as it was currently implemented in most states did not contain the critical elements necessary to have a positive impact on student achievement (Hill, 2009).

Cameron (2011) studied the impact of professional growth plans on student achievement in the Kennewick School District in Washington. The school district had been implementing a program called Professional Growth Plans for Clock Hours for four years in which teacher volunteers could use the hours devoted to implementing the plans in place of the required time spent in professional development workshops to gain hours

for certificate renewal. In comparing the Washington Assessment of Student Learning scores of schools with a 33% professional growth plan participation rate to those with a 4% participation rate, no effect on the passing rates was noted.

### *Qualitative Outcomes*

In surveying teachers on individualized growth plans, Peterson, Wahlquist, Bone, Thompson, and Chatterton (2001) found that morale improved with 82.5% of the teachers enjoying the growth plans due to the increased control of their learning. Peterson et al. also concluded that individual growth plans led to increased reflection about the practice of teaching as compared to generic professional development opportunities.

Dunaway, et al. (2010) conducted a qualitative study on individual growth plans in the Iredell-Statesville School system in North Carolina where they had been implementing them for three years. Both teachers and administrators responded to a survey instrument with a 5-point Likert scale focusing on the process, purpose, and the value of individual growth plans. The results indicated that the teachers thought that an appropriate measurement of the plan was improved performance, and the attainment of the goals should be included in their summative evaluations. The teachers identified a fault of the system as being its fall-to-spring cycle instead of a spring-to-spring cycle. The survey results also indicated that teachers did not see the individual growth plan as having a significant impact on the learning outcomes of students and school-wide performance. The principals' overall scores were higher than the teachers' indicating that they took the process more seriously and saw it as more collaborative in nature. Dunaway et al. (2010) concluded that further research is needed on a larger scale.

According to Dunaway et al. professional growth at the individual level is necessary for organizational growth to occur.

### *Summary*

The link between student achievement and professional development of teachers is still not clear. Although there is a lot of research focused on practitioner inquiry at the pre-service level as well as descriptive procedural findings, the empirical research with regards to the consequences of practitioner inquiry on student learning are limited. Studies that seek to discover a connection between a teacher's professional growth and student achievement are typically small scale, and when they are conducted on a larger scale, they are usually qualitative based on self-evaluations and teacher opinions. The review of literature resulted in few studies that linked professional development with student academic outcomes and no studies were located that directly determined the relationship between the job-embedded professional development format of individualized professional growth plans and student academic outcomes.

### Collaboration and Mutual Accountability

#### *Definition*

Although there were several models including the components of collaboration and mutual accountability, the most prominent model in literature was what DuFour (2004) called professional learning communities (PLCs). Bolam et al. (2005) defined a professional learning community as "A community with the capacity to promote and sustain the learning of all professionals in the school community with the collective purpose of enhancing student learning" (p. 145). DuFour (2004) further emphasized

teacher collaboration as a critical component of this process as well. In reviewing studies relative to collaboration and mutual accountability among teachers, professional learning communities will be used synonymously.

### *Teacher Outcomes*

According to Musanti and Pence (2010), extensive qualitative research has been conducted focusing on collaboration across all disciplines. More specifically in education, research has concluded that professional development aimed at developing collaboration positively improves teaching effectiveness (Berry, Daughtrey, & Wieder, 2009; Fulton & Britton, 2011; Musanti & Pence, 2010). Musanti and Pence (2010) found that collaboration furthered teachers' opportunities to develop networks through which practices could be reflectively shared, teaching and learning beliefs could be discussed, and knowledge could be co-constructed. Reform efforts most recently have emphasized collaboration among teachers (Brownell, Yeager, Rennells, & Riley, 1997; Louis, Marks, & Kruse, 1996).

Brownell et al. (1997) in their review of research on collaboration in educational settings found that increased teacher collaboration was associated with greater efficacy, improved affect, and an enhanced knowledge base. Because research has shown that increased self-efficacy among teachers was associated with increased student achievement, it was concluded that teacher collaboration had a positive impact on student achievement (Englert, Tarrant, & Rozendal, 1993; Ross, 1992). Smylie, Lazarus, and Brownlee-Conyers (1996) in studying teacher control and autonomy found a negative relationship between teacher autonomy and student achievement. They therefore

concluded that collaboration including mutual accountability would improve student learning.

In a qualitative study of 1,213 Tennessee teachers in 78 elementary schools, Rosenholtz and Simpson (1990) concluded that teachers with experience were satisfied from collaborating with colleagues. They reported that sharing ideas to improve student performance increased their commitment. Katzenbach and Smith (1993) concluded from studying teams from 47 organizations that they outperformed teachers who worked in isolation. Additionally, when teachers collaborated about instruction, they held higher student and teacher expectations, they were more innovative, and they demonstrated a stronger commitment to teaching. According to Goddard, Goddard, and Tschannen-Moran (2007), although collaboration enabled teachers to enhance their experiential and pedagogical knowledge contributing to improved instruction and therefore student achievement, this remained to be only theoretical until further research was conducted that directly linked student outcomes with collaboration.

### *Student Outcomes*

According to Schmoker (2004), collaboration among teachers has a significant effect on student performance. Several studies had been conducted to determine the impact of collaboration on student outcomes. Supovitz (2002) studied reform efforts in the Cincinnati Public Schools based on small learning communities. This study looked at reform involving teams of teachers in Grades 4-8 who worked collaboratively to develop instructional strategies for improving the learning of the students in writing, reading, mathematics, science, and citizenship for whom they were accountable. The teachers through survey responses reported more collaboration and interactions with their

colleagues as a result of the reform efforts; however, collective instructional practices were limited with 70% of the respondents indicating low usage, 25% of the respondents indicating moderate usage, and only 5% of the respondents indicating high usage. Teacher groups spent more time collaborating on preparation and grouping strategies than on instructional practices. The study concluded that there was a positive relationship that was statistically significant between collective instructional practices and student achievement in 14 of the 25 grade level subject areas studied. The correlations ranged from .05 in fifth grade citizenship to .23 in fourth grade writing. Supovitz concluded that professional learning communities that focus on discussions centered on the relationship between instructional strategies and student performance result in significant student learning gains. He found that for every .10 standard deviation a team incorporated collective instructional practices, test scores would increase by .10 standard deviations.

Phillips (2003) conducted a case study over a three-year period in a Texas urban middle school in which the teachers collaboratively targeted students who were underachieving in reading, mathematics, science, writing, and social studies. As a result, student achievement on the Texas Assessment of Academic Skills (TAAS) went from 50% of the students passing all subject area tests in 1999-2000 to 90% of the students passing all subject area tests in 2001-2002.

Strahan found similar results in 2003 when he studied the third through fifth grade students in three struggling North Carolina elementary schools with an average minority rate of 77% and an average free and reduced lunch rate of 74%. In 1997 all three schools had less than half of the students score on grade level or above according to the North Carolina end-of-grade assessments. After five years of implementing a collaborative



culture in which teachers shared accountability for students, in 2002 the average percentage of students on or above grade level at all three schools rose to an average percentage of 75.6.

Hollins, McIntyre, DeBose, Hollins, and Towner (2004) studied a California elementary school in an urban school district with 300 Grade K-5 students and an African American population of 91% to determine the impact of collaboration on student achievement. Collaboration consisted of identifying and implementing new practices through a collective team approach in order to meet student needs. The Stanford Achievement Test was used to compare baseline scores to scores over a two-year period. In reading at the second grade level there was an increase of 28% of the students scoring above the 25 percentile as compared to a 12% increase at the district. In reading at the third grade level there was an increase of 31% of the students scoring above the 25 percentile as compared to 9% at the district.

Gruenert (2005) investigated the relationship between the collaborative culture of a school and student achievement. He surveyed 2,750 elementary, middle, and high school teachers from 81 Indiana schools. The survey instrument provided a score based on 35 collaborative item descriptors grouped into six factors. These factors were each correlated with mathematics and language art Indiana Statewide Testing for Educational Progress (ISTEP) scores. Findings indicated that schools that had a more collaborative culture at all levels had higher ISTEP scores on average. The strongest correlations were associated with teachers viewing themselves as learners ( $r = .278$  for mathematics and  $r = .234$  for language arts), a clear and unifying mission statement ( $r = .455$  for mathematics

and  $r = .397$  for language arts), and parent involvement ( $r = .471$  for mathematics and  $r = .506$  for language arts).

Goddard et al. (2007) conducted a study in an urban, midwestern school district using a random sample of 452 elementary teachers across 47 schools who taught 2,536 fourth grade students. The level of collaboration among teachers was measured using a survey and was then correlated to student assessment data using a state standardized assessment to determine if collaborative practices among teachers aimed at school improvement were related to student achievement. The results indicated that students' mathematics and reading performance were better at schools that had a higher level of collaboration among its teachers. More specifically, an increase of one standard deviation in the level of collaboration among teachers was associated with an increase of .08 standard deviations in a school's average achievement in mathematics and .07 standard deviations in a school's average achievement in reading.

Jackson and Bruegmann (2009) studied third through fifth grade teachers from 1995-2006 in North Carolina to determine the collective effect of same grade level teacher peers on the achievement of students on standardized tests. The study concluded that gains on test scores were greater for students whose teachers had peer teachers with higher value-added mathematics and reading mean estimates. Results further indicated that an improvement of one standard deviation in the estimated quality of a peer teacher was associated with an increase in reading scores of .026 and in mathematics scores of .0398. Jackson and Bruegmann further concluded that 20% of value-added teacher effects as measured by student test gains could be attributed to collaborative expertise.

The Teachers Network, a national nonprofit organization, carried out a national survey involving 1,210 teacher leaders to determine the role collaboration played in their effectiveness (Berry et al., 2009). As a result of collaboration, 90% of the teachers thought that it contributed to the improvement of their teaching, and over 75% thought that it contributed to improvement of their school overall.

### *Summary*

Overall the research on collaboration aimed at improving instructional practice confirms its positive effect on student achievement. Providing teachers with ongoing opportunities to share information about students, instruction, and challenges results in a mutual responsibility focused on gains. Although the research indicated a relationship between teacher collaboration and student achievement, it did not provide conclusive evidence of a cause-effect relationship due to the nature of the studies.

### Value-Added Assessment

#### *Definition*

“A value-added model is a statistical model that uses student-level growth scores to differentiate teacher performance in the area of student learning growth” (American Institute for Research, 2011, slide 14). According to the American Institute for Research, unique factors associated with students and schools are separated from those factors associated with a teacher. Value-added models consider students’ test scores both current and historically to estimate the effect that a teacher had on their learning growth (Hill et al., 2010). Student learning based on past test scores is used to predict future performance (American Institute for Research, 2011). The difference between a

student's predicted performance and his actual performance signifies the value added by the classroom teacher's instruction (American Institute for Research, 2011). According to Hill et al. (2010), the value-added model has grown in popularity because of the evidence that suggests that teachers influence their students' learning growth. In fact, Nye, Konstantopoulos, and Hedges (2004) found that teacher effect accounted for 11% of the variation in student assessment gains. Gordon (2008) further found that the best predictor of teachers' value-added performance was their value-added scores from the previous year.

### *Models*

Hill et al. (2010) conducted a survey of school districts across the United States that used value-added scores for accountability purposes and found that there was little agreement as to the value-added model specifications. A variety of value-added models exist and are being used that differ on the covariates being controlled as well as the number of previous year data included in the formula (Hill et al., 2010). Gordon (2008) conducted a study to determine the impact of controlling for student background characteristics when determining the effect of a teacher on student achievement. Student test scores in mathematics were used to determine teacher value-added scores when only controlling for student baseline scores. Then further controls including race, ethnicity, gender, socioeconomic status, and language were controlled and the value-added scores recalculated. Gordon found that the correlation between the value-added scores was 0.98 and therefore concluded that as long as baseline test scores were being controlled; models that included additional controls only made modest differences in value-added scores.

According to Hill et al. (2010), the most widely used model, the Education Value Added Assessment System (EVAAS) Multiple Response Model (MRM), does not control for student nor school effects but does control for district effects in the state model by using scaled scores from statewide distributions. Based on 24 years of research, Sanders (2006) argued the merits of this model as being the most robust and conservative.

Newton, Darling-Hammond, Haertel, and Thomas (2010) conducted a study of 250 San Francisco Bay Area secondary mathematics and English language arts teachers to determine the stability of teacher effectiveness value-added ratings across years, classes or courses taught, and statistical models based on the California Standards Tests. Five models were used to calculate teacher effect, four of which were included in this summary because of their popularity. Results were based on models using only prior achievement, prior achievement and student characteristics, prior achievement and school fixed-effects, and prior achievement along with student characteristics and school fixed-effects. Results indicated that the teacher ratings among the four models were correlated at a high level between .83 and .93. Newton et al. (2010) concluded that the differences in teacher rankings were significantly related to student demographics which in this study included ethnicity, gender, parents' level of education, and socioeconomic status. When the same teachers taught one class of higher-track students one year and another class of lower-achieving students the other year, their ranking was significantly higher for the higher-track students. In fact, teachers ranked in the 7<sup>th</sup> to 9<sup>th</sup> deciles with higher-track students compared to the 1<sup>st</sup> to 3<sup>rd</sup> deciles with lower-achieving students. Newton et al. (2010) further cautioned using value-added measures due to compositional or contextual

effects. That is, students' achievement was not only affected by individual characteristics but also by the other students' characteristics in the class. Because of the increasing prominence of value-added assessment relative to the effectiveness of teachers, there is an increased amount of research being conducted exploring the sensitivity of estimates based on various models (Papay, 2011).

### *Reliability*

Koedel and Betts (2007) conducted a study that examined the effects of San Diego elementary school teachers' value-added scores in reading and mathematics to student performance on the Stanford 9 standardized exam. The sample was comprised of 16,000 students in Grades 2-5 and 1,000 teachers. Koedel and Betts concluded that a change in teacher quality by one standard deviation in mathematics resulted in a 0.41 change in student years and an average test score gain of 0.26; whereas a change in teacher quality by one standard deviation in reading resulted in a 0.31 change in student years and an average test score gain of 0.19. Koedel and Betts also found that teacher qualifications that were observable were weakly related to teacher quality as measured by value-added scores, and only 0.9 to 1.4% of the variance in teacher quality could be explained by compensation. When examining the consistency of value-added scores over time, Koedel and Betts found that the teachers in the top and bottom quintiles were more likely to retain their positions as compared to those teachers in the middle. Koedel and Betts further cautioned the reader that high levels of estimation errors resulted in relatively low reliability with the variance decomposition indicating teacher quality variance at 60% of the total fixed-effects variance in mathematics and 50% of the total fixed-effects variance in reading. Additional research conducted on value-added scores

also indicated that they were made up of equal amounts of variance contributed to error and real score (Lockwood, Louis, & McCaffrey, 2002; McCaffrey, Sass, Lockwood, & Mihaly, 2009).

An executive summary of The Economic Policy Institute concluded that researchers who had analyzed VAM results found that the methodology was not reliable in identifying effective and non-effective teachers (Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn, Ravitch, Rothstein, Shavelson, & Shepard, 2010). Several inclusive studies had shown that a teacher's effectiveness according to VAM fluctuates over time, statistical models, and classes. According to the Board on Testing and Assessment of the National Research Council of the National Academy of Sciences, teacher effectiveness based on VAM scores are not reliable enough to be considered in making operational decisions (Baker et al., 2010). According to Baker et al. the fact that students are not assigned to teachers randomly and their home and school experiences vary, means that even if student characteristics are controlled in statistical models, they cannot be used to accurately judge teacher effectiveness. Baker et al. based this on the fact that education is cumulative, and a single teacher does not account for the total achievement of a student. Ishii and Rivkin (2009) also concluded that family heterogeneity is difficult to control, and not accounting for differences among families impacts measurements of teacher quality unless classrooms are assigned randomly when making student placement decisions. Even when students are randomly assigned to teachers, Ishii and Rivkin pointed out how school and family efforts to compensate for poor teacher quality; such as, tutoring and additional parent support can mask a poor-quality teacher's value-added score.

McCaffrey, Lockwood, Koretz, Louis, & Hamilton (2004) questioned the reliability of VAM results due to the sampling error associated with small class sizes especially at the elementary level and the compounded measurement error associated with multiple years of testing. Baker et al. (2010) found an error rate of 26% when using three years of data and an error rate of 36% when one year of data was available. This means that one out of four teachers when using three years of data and one out of three teachers when using one year of data could be misclassified as effective or not effective based on VAM results. Additionally, Koretz (2008), a testing expert, found that in order to use value-added models to measure growth across grade levels, the test should be vertically scaled, or measure test content from year to year along a continuum. Koretz found that many standard based assessments used to determine VAM are not vertically scaled.

A Measures of Effective Teaching study funded by the Gates Foundation reported reliability for VAM to be 0.3 to 0.5 when based on data from three years (Harris, 2012). This means that of the teachers who ranked in the top quintile according to their VAM one year, only 28 to 50 percent were ranked similarly in subsequent years, and 4 to 5 percent moved from the top quintile to the bottom quintile.

### *Validity*

Hill et al. (2010) studied the congruence of survey, observation, and value-added data related to teacher quality of 24 middle school mathematics teachers within a southwestern school district. The instruments used in the study included the state mathematics assessment, a mathematics knowledge test composed of 159 survey items, and a mathematical quality of instruction observational instrument. Hill et al. concluded



that despite controlling for past performance, students past scores were strong predictors of teachers' value-added scores with an  $r$  value of 0.27 to 0.53. In other words, teachers with higher level students had, on average, higher value-added scores. Sanders (2006) argued that since measurement error biases covariate adjustment models, using previous tests scores from at least three years minimizes this problem.

Braum (2005) noted that the value-added scores were more unstable for teachers at the lower and upper ends of the scale. A study conducted by Sass (2008) of five larger urban school districts found that of the teachers ranked according to their value-added scores in the bottom 20% on effectiveness, less than one third were ranked in the bottom 20% the second year while another third was ranked in the top 40%. Likewise, among the teachers ranked in the top 20% in the first year, in the second year only one third remained in the top 20%. Another study conducted by McCaffrey et al. (2009) found that year-to-year correlations of teacher quality ranged from 0.2 to .4. Based on their studies and review of research of VAM, Baker et al. (2010) concluded that "There is not a stable construct measured by value-added measures that can readily be called teacher effectiveness" (p. 13).

### *Outcome Measures*

Using six years of longitudinal data, Papay (2011) compared the value-added scores for teachers based on two different mathematics Stanford Achievement Test (SAT) subscales. Papay concluded that the choice of outcome measure had a greater impact on the value-added score than the model used. Papay further studied the phenomenon by comparing the value-added score for teachers based on three different reading achievement test scores; the SAT, the state assessment, and the Scholastic

Reading Inventory. The results indicated a statistically significant correlation between the value-added scores derived from the three tests. Therefore, Papay concluded that teachers whose students perform well on one assessment score similarly on similar assessments. Papay further concluded that although there was a moderate correlation between the subsequent value-added scores, the rank of teachers based on the scores differed considerably. According to Papay, “If this district implemented a high-stakes pay-for-performance program similar to the one currently operating in Houston, Texas, simply switching the outcome measure would affect the performance bonuses for nearly half of all teachers and the average teacher’s salary would change by more than \$2,000” (p. 165).

Corcoran, Jennings, and Beveridge (2011) further supported Papay’s (2011) findings about the impact of different tests on value-added measures with the results from a study they conducted with fourth and fifth grade reading and mathematics students in Houston. The Houston Independent School District administered two standardized tests; the Stanford Achievement Test (SAT), a low-stakes test used for diagnostic purposes, and the Texas Assessment of Knowledge and Skills (TAKS) test, a high-stakes test used to reward and punish teachers. Using these two tests, Corcoran et al. (2011) conducted a study to determine if value-added ratings differed based on the test administered. Results indicated that teacher effect was magnified on high-stake tests by 15-31% resulting in inconsistent teacher ratings across tests. A teacher’s value-added estimate on the SAT as compared to the TAKS was only modestly correlated in both reading,  $r = .499$  and mathematics  $r = .587$ . Therefore, value-added on the low-stakes test was not a predictor of value-added on the high stakes test. Corcoran et al. concluded that there was a

stronger correlation in teacher effects on the same test between subjects than on different tests in the same subject area. Corcoran et al. concluded that based on these correlations, a highly effective teacher as identified by student scores on a state test might not be highly effective on a low-stakes test even though the subject, time of year, and set of students were the same. Quintile rankings also varied across tests with only 43% of the teachers ranked in the top quintile on the SAT also ranked in the top quintile on the TAKS.

### *Florida's Model*

Florida's value-added model is a covariate adjustment model (Florida Department of Education, 2011). In other words, statistical methods are used to compensate for an imbalance caused by the influence of variables that cannot be controlled. Based on up to three years of student data, a predicted performance value is determined for each student after accounting statistically for student, classroom, and school characteristics using a value-added model. The following are the student- and classroom-level characteristics being controlled as identified by the Florida Department of Education (2011):

- up to two prior years of achievement scores,
- the number of subject-relevant courses in which the student is enrolled that are linked to an FCAT test by course code,
- Students with Disabilities (SWD) status receiving special education services,
- English Language Learner (ELL) status enrolled in a program for less than two years,
- gifted status,
- attendance, or number of days present,
- mobility, or number of transitions within the school year,
- difference from modal age of students in that grade across the state,
- class size, or the number of students linked to a teacher, and
- homogeneity of students' entering test scores in the class as a continuous variable based on the interquartile range. (p. 14)

To control for school-level characteristics, a school component is also calculated. It is based on the average student performance in a school that differs from the statistical expectation. In calculating the VAM score for a teacher, 50% of the school component will be attributed to the teacher. According to the Florida Department of Education (2011), “One recognizes that the teacher contributes somewhat to the overall school component, but there are factors imbedded in that component that are beyond his/her direct control and that he/she should not directly be held accountable for” (p. 16). Once a student’s predicted performance is calculated, the difference between the actual performance and the predicted performance is the value-added by the teacher.

#### *Summary*

The reliability of VAM scores was questioned in the literature based on levels of estimation error, sampling error, and the inability to control for variables that impact assessment results beyond school and teacher effects. Likewise, researchers challenged the validity of VAM scores. Research confirmed that VAM scores fluctuated from year to year, especially for those teachers in the bottom quartile rating, and therefore, the stability of the construct being measured was likewise questioned. Although various value-added models exist, according to research, the most important covariates to control to assure a strong correlation between the value-added score and student achievement are baseline scores, student demographics, and outcome measures.

## CHAPTER III METHODOLOGY

The purpose of this study was to determine the relationship between the professional practices components of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument and student achievement as measured by a teacher's value-added measurement score. With the 2011-2012 school year being the initial year of implementation for the instrument, the problem existed of needing to confirm the quality of the instrument as determined by its construct validity to assure that it served the intended purpose of measuring quality instruction. This chapter contains a detailed explanation of the design of the research that was conducted, a description of the population, the procedures used to conduct the research, the instrumentation used to gather the data, and data analysis procedures relative to the research questions and hypotheses for the study. Permission was granted by Dr. Debra Pace on April 1, 2012 to reproduce all Brevard County Public School evaluation figures and instruments included in this chapter (see Appendix D for Permission to Reproduce).

### Research Design

This quantitative, correlational research study was designed to test the extent to which there was a relationship between the professional practices component ratings and the value-added assessment score on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. The extent to which relationships exist was not to suggest causal inference.

## Population

This study's population was comprised of all Brevard Public School fulltime instructional personnel during the 2011-2012 school year in Grades 4-10 in which reading and/or mathematics was taught, FCAT was administered, and they received a teacher aggregated VAM score as opposed to a school aggregated VAM score or a combination thereof. A list of the specific courses included in this study is located in Appendix E. Instructional personnel received a teacher aggregated VAM score if (a) they were hired as a first-year teacher for the 2011-2012 school year, (b) they had a combination of two VAM estimate scores for any two years, or (c) they had three years of VAM estimate scores (Brevard Public Schools, 2011). The population excluded the instructional personnel at charter schools, the adult education centers, alternative learning centers, virtual schools, and the school board department. The total number of instructional personnel who met the criteria for participation was 1,138. This population was comprised of 736 (64.68%) elementary school teachers, 177 (15.55%) middle school teachers, 59 (5.18%) junior/senior high school teachers, and 166 (14.59%) high school teachers. The gender of the population was 984 (86.47%) females and 154 (13.53%) males. The ethnicity of the population was 3 (.26%) American Indian or Alaskan, 6 (.53%) Asian or Pacific Islander, 51 (4.48%) Hispanic, 70 (6.15%) Black, and 1,008 (88.58%) White, Non-Hispanic. Years experience of the population was 505 (44.38%) 0-10 years experience, 413 (36.29%) 11-20 years experience, 159 (13.97%) 21-30 years experience, 59 (5.18%) 31-40 years experience, and 2 (.18%) 41 and higher years experience.

## Methods of Data Collection

The data for this study were based on 2011-2012 Brevard Public Schools instructional personnel evaluation scores accessible with permission from the Human Resource Department where every teacher's Evaluation Summary Calculation Form was filed (see Appendix F for the Evaluation Summary Calculation Form). The Evaluation Summary Calculation Form contains all of the professional practices ratings used for this study. It consists of four subscales including the Formative Evaluation Observation Component and the Continuous Professional Improvement Component that were the focus of this study. The Formative Evaluation Observation Component measures instructional practices, and the Continuous Professional Improvement Component measures professional development. The Formative Evaluation Observation Component consists of two subscales including the Quality of Instruction dimensions and the Professional Responsibility, Conduct and Relationships dimensions. The Continuous Professional Improvement Component consists of three subscales including Professional Growth Plan Development, Professional Growth Plan Implementation, and Collaboration and Mutual Accountability. Although this form contains additional data, it was not needed for this study. Aggregated teacher VAM scores were provided by the State Department of Education following the release of the 2012 FCAT scores to the school district for all teachers of reading and mathematics who administered the FCAT.

A list of all instructional personnel who met the parameters for the population was obtained from the school district human resource database. Demographic information for the instructional personnel along with their data were matched by employee number, then de-identified with a dummy code, and finally compiled into Microsoft Excel. For

statistical analysis, these data were entered into the software program The Statistical Package for Social Sciences (SPSS). Approval for this study was submitted to The University of Central Florida Institutional Review Board and the Brevard Public Schools (see Appendix G for the IRB Review and Brevard Public Schools Research Approval).

### Instrumentation

The School Board of Brevard County Instructional Personnel Performance Appraisal Instrument is divided into two separate components; Professional Practices and Student Performance (Brevard Public Schools, 2011). The Professional Practices component is comprised of a possible 21 points based on seven dimensions of Florida's Accomplished Practices, a possible 10 points for the development of a Professional Growth Plan, a possible 8 points for the implementation of a Professional Growth Plan, and a possible 8 points for working collaboratively with colleagues and being mutually accountable for student performance. The Student Performance component of the instrument is comprised of a possible 35 points for the value-added growth measure, a possible 5 points for collaborative team effort in closing the achievement gap of the lowest 25% students in reading and/or mathematics, a possible 5 points for the implementation and achievement of School Improvement Plan goals, a possible 2 points for achieving the Professional Growth Plan targets, and a possible 3 points for student achievement in science and overall as regressed against similar demographic school data within the state of Florida. In addition to the 97 possible points inclusive of these two components, a possible 3 points are awarded for the alignment of the professional practices with student growth measures. See Figure 3 for a visual representation of the Performance Appraisal Model as outlined above.



<b>50% Professional Practices</b>	<b><u>PROFESSIONAL PRACTICES BASED ON FLORIDA'S EDUCATOR ACCOMPLISHED PRACTICES</u></b>	<b>50%</b>	<b>50% Student Performance</b>	<b><u>INDIVIDUAL ACCOUNTABILITY FOR STUDENT GROWTH BASED ON IDENTIFIED ASSESSMENTS</u></b>	<b>40%</b>
	21 points: Professional Practices			35 points: Individual accountability for <b>student achievement/value added growth measures</b>	
	10 points: Professional Growth Plan Development			5 points: Collaborative team student achievement results related to <b>closing the achievement gap</b> of the Lowest 25% in Reading and/or Math using value-added measures.	
	8 points: Plan Implementation			<b><u>DISTRICT OPTION: TEAM AND SCHOOL ACCOUNTABILITY</u></b>	<b>10%</b>
	8 points: Collaboration/Mutual Accountability			5 points : Achievement of <b>School Improvement Plan goals</b> assigned for whole school results <u>or</u> team results	
3 points: Alignment of Professional Practices with Student Performance Measures	2 points: Individual accountability for meeting individual <b>Professional Growth Plan (PGP) target(s)</b>				
				3 points: Whole School student achievement results as measured by <b>regression</b> of FCAT components of School Grade	

Figure 3: Performance Appraisal Model (Brevard Public Schools, 2011, p.17).

The points for the Professional Practices components of the instrument are recorded onto an Evaluation Summary Calculation Form which was used to gather the data for this study (see Appendix F for the Evaluation Summary Calculation Form). Following is a detailed description of the instruments used to determine the data for the Evaluation Summary Calculation Form. Since these instruments were designed specifically for Brevard County Public Schools, the validity of the instruments was based on the research on which they were designed. No previous evidence for the validity of the instruments existed beyond their research base because they were not field tested prior to the 2011-2012 school year implementation. Several procedures have been planned and executed to assure the reliability of the instrumentation.

### *Reliability*

Inter-rater reliability is the consistency of results by observers using the same instrument (Lomax, 2007). According to Thornton (2012), inter-rater reliability of an appraisal instrument will improve if observers are trained on the method of objective

rating and the rating scales are specific and clearly defined. According to Dori Bisbey, Director of Educational Leadership and Professional Development, in order to improve the inter-rater reliability of the Brevard County Instructional Personnel Performance Appraisal system, training has been ongoing (personal communication, April 10, 2012).

Seven sessions were offered in June and July 2011 to provide an overview of the new instructional personnel performance appraisal system to all leadership team members along with three teacher leaders from every school. Four sessions were offered in August 2011 to train assistant principals and district review teams on the system. Three professional growth plan trainings were offered in August and September 2011, one in each area of the school district, for administrators and the three teachers selected from each school to be peer review team members on how to use the rubric to score the professional growth plans. In September 2011, the principals were provided a full day of formative practice on scoring teachers using the observation rubrics, and in October 2011 the assistant principals were provided the same full day training along with the three additional training sessions. In November 2011, the principals were given a test for inter-rater reliability on using the observation rubrics, and in December 2011 and January 2012 assistant principals were given the same test. This test required administrators to view a short teaching video and score the teacher using the observation rubrics. The results for the principals ( $n = 110$ ) indicated a mean score of 38.96 out of a possible 63 points with a standard deviation of 5.74. The scores ranged from 21 to 52 with a median score of 40. The results for assistant principals ( $n = 66$ ) indicated a mean score of 40.86 out of a possible 63 points with a standard deviation of 4.91. The scores ranged from 29 to 54 with a median score of 41. In January and February 2012, six sessions were offered to

administrators and secretaries on how to complete identified forms required by schools for the evaluation. Twenty one-day sessions entitled “Analyzing Rubrics for Effective Teaching” have been offered to teacher leaders focused on interpreting the observation rubrics and distinguishing between the descriptors associated with the ratings. Six two-day coaching trainings were conducted for teams of three teachers from each school designed to assist teachers as they coach their peers. All teachers who attended this training were required to have clinical educator training (CET) as a prerequisite; therefore, nine CET trainings were offered. Eight supervisory coaching trainings were conducted for assistant principals to train them on how to effectively conference with teachers. All principals were also required to receive a one-day initial training in coaching and mentoring. In addition to the scheduled trainings as outlined above, ongoing practice involving viewing teaching videos and scoring them using the observation rubrics has been conducted at monthly leadership and assistant principal meetings. Also, it has been encouraged that all district level training be duplicated at the school level.

### *Classroom Observation Instrument*

The purpose of the Classroom Observation Instrument is to collect data through the examination of teaching in the classroom in order to provide teachers with feedback on which to reflect and improve instructional practices (Brevard Public Schools, 2011) (see Appendix H for the Classroom Observation Instrument). In addition to feedback provided to teachers through informal classroom walkthroughs, teachers with zero to three years of teaching experience are required to have two formal observations by an

administrator, and teachers with four or more years of experience are required to have one formal observation by an administrator.

The Classroom Observation Instrument measures the level of competency of teachers in four observable dimensions including Learning Environment, Instructional Delivery and Facilitation, Assessment, and Relationship with Students. The Learning Environment dimension consists of eight subscales, or elements, including items such as “Manages student conduct” and “Maintains a climate of inquiry” (Brevard Public Schools, 2011, Appendix 24). The Instructional Delivery and Facilitation dimension consists of six subscales, or elements, including items such as “Employs higher order questions” and “Differentiates instruction” (Brevard Public Schools, 2011, Appendix 25). The Assessment dimension consists of four subscales, or elements, including items such as “Modifies teacher made assessments to accommodate diversity” and “Communicates assessment data to students and parents” (Brevard Public Schools, 2011, Appendix 26). The Relationship with Students dimension consists of three subscales, or elements, including items such as “Demonstrates knowledge of students” and “Creates a positive environment of respect and rapport” (Brevard Public Schools, 2011, Appendix 27).

For each subscale, or element, a 4-point Likert scale ranging from unsatisfactory to distinguished is to be determined. The formative scale for observations is as follows:

- Distinguished (3 pts): Indicates performance that consistently exceeds the requirements of the position and the level of performance commensurate with the experience of the teacher,
- Proficient (2 pts): Indicates performance that consistently meets the requirements of the position and the level of performance commensurate with the experience of the teacher,
- Professional Support Needed (1 pt): Indicates performance that requires additional attention to ensure an accepted level of proficiency. Further, this performance is not characteristic of the requirements for the position and experience of the teacher,

- Unsatisfactory (0 pts): Indicates performance that does not meet the minimum requirements of the position and the level of performance commensurate with the experience of the teacher (Brevard Schools, 2011, p. 20).

There is also a comment section for each subscale, or element, along with a place to record the date and time of the observation.

### *Observational Rubrics*

The rubrics used for both formative and summative evaluation purposes are based on the observable Florida Educator Accomplished Practices, or what students seeking to become certified teachers by state-approved education programs must know, and Brevard Public School Standards based on research (Brevard Public Schools, 2011) (see Appendix I for the Observational Rubrics). The rubrics provide a means of measuring teacher competency as defined by seven dimensions including Instructional Design and Lesson Planning, Learning Environment, Instructional Delivery and Facilitation, Assessment, Professional Responsibilities and Ethical Conduct, Relationship with Students, and Relationship with Parents and Community.

#### *Instructional Design and Lesson Planning*

The Instructional Design and Lesson Planning dimension consists of four subscales, or elements, including “Sets instructional outcomes and aligns instruction with state-adopted standards” (Brevard Public Schools, 2011, Appendix 5). A 4-point Likert scale ranging from unsatisfactory to distinguished is provided for each subscale, or element, as well as a description of specific teacher behaviors that must be evident in order to obtain the rating. For example, in order to receive a rating of proficient for the element, “Sets instructional outcomes and aligns instruction with state-adopted

standards,” a teacher must demonstrate that “Lesson plans are aligned to the district adopted curriculum maps and district/state assessments” (Brevard Public Schools, 2011, Appendix 5).

### *Learning Environment*

The Learning Environment dimension consists of eight subscales, or elements, including “Maintains a climate of inquiry” (Brevard Public Schools, 2011, Appendix 8). A 4-point Likert scale ranging from unsatisfactory to distinguished is provided for each subscale, or element, as well as a description of specific teacher behaviors that must be evident in order to obtain the rating. For example, in order to receive a rating of proficient for the element, “Maintains a climate of inquiry,” a teacher must demonstrate the following: “Engages students in problem solving inquiry-based activities through the use of high level questioning” and “Student participation and responses indicate individual understanding of content and/or concepts” (Brevard Public Schools, 2011, Appendix 8).

### *Instructional Delivery & Facilitation*

The Instructional Delivery & Facilitation dimension consists of six subscales, or elements, including “Differentiates instruction” (Brevard Public Schools, 2011, Appendix 10). A 4-point Likert scale ranging from unsatisfactory to distinguished is provided for each subscale, or element, as well as a description of specific teacher behaviors that must be evident in order to obtain the rating. For example, in order to receive a rating of proficient for the element, “Differentiates instruction,” a teacher must demonstrate “Evidence of incorporating various differentiated instructional strategies (e.g. ability

grouping or compacting of lessons) to meet the needs of students with varying learning styles and abilities” (Brevard Public Schools, 2011, Appendix 10).

### *Assessment*

The Assessment dimension consists of four subscales, or elements, including “Modifies teacher made assessments to accommodate diversity” (Brevard Public Schools, 2011, Appendix 12). A 4-point Likert scale ranging from unsatisfactory to distinguished is provided for each subscale, or element, as well as a description of specific teacher behaviors that must be evident in order to obtain the rating. For example, in order to receive a rating of proficient for the element, “Modifies teacher made assessments to accommodate diversity,” a teacher must demonstrate the following: “Appropriately differentiates assessments to address the unique learning differences of students that have a wide range of learning styles and abilities” and “Provides a variety of assessments to meet the needs of students” and “Students may have choices in their assessments” (Brevard Public Schools, 2011, Appendix 12).

### *Professional Responsibilities & Ethical Conduct*

The Professional Responsibilities & Ethical Conduct dimension consists of five subscales, or elements, including “Applies technology to organize and communicate assessment information” (Brevard Public Schools, 2011, Appendix 14). A 4-point Likert scale ranging from unsatisfactory to distinguished is provided for each subscale, or element, as well as a description of specific teacher behaviors that must be evident in order to obtain the rating. For example, in order to receive a rating of proficient for the element, “Applies technology to organize and communicate assessment information,” a

teacher must demonstrate “Using technology to communicate student learning and assessment information to appropriate stakeholders in a timely manner” (Brevard Public Schools, 2011, Appendix 14).

### *Relationship with Students*

The Relationship with Students dimension consists of three subscales, or elements, including “Builds relationships through instructional interactions” (Brevard Public Schools, 2011, Appendix 15). A 4-point Likert scale ranging from unsatisfactory to distinguished is provided for each subscale, or element, as well as a description of specific teacher behaviors that must be evident in order to obtain the rating. For example, in order to receive a rating of proficient for the element, “Builds relationships through instructional interactions,” a teacher must demonstrate the following: “Teacher allows and encourages most students to be part of class discussions and interactions” and “Teacher brings some student interests into the content” (Brevard Public Schools, 2011, Appendix 15).

### *Relationship with Parents and Community*

The Relationship with Parents and Community dimension consists of ten subscales, or elements, including “Maintains a family friendly learning environment” (Brevard Public Schools, 2011, Appendix 17). A 4-point Likert scale ranging from unsatisfactory to distinguished is provided for each subscale, or element, as well as a description of specific teacher behaviors that must be evident in order to obtain the rating. For example, in order to receive a rating of proficient for the element, “Maintains a family friendly learning environment,” a teacher must demonstrate “Maintaining a family



friendly environment that encourages engagement (inviting climate, opportunities to volunteer in the classroom, and encourage collaboration with parents)” (Brevard Public Schools, 2011, Appendix 17).

### *BPS Instructional Personnel Evaluation Instrument*

The BPS Instructional Personnel Evaluation Instrument is used for summative purposes and measures a teacher’s overall quality of instructional performance as defined by the Observational Rubrics (Brevard Public Schools, 2011) (see Appendix J for the BPS Instructional Personnel Evaluation Instrument). The top of the instrument contains demographic information including the teacher’s name, school year, school name, school number, contract status, principal or department head name, assignment, and employment status. The instrument consists of seven performance areas including Instructional Design and Lesson Planning, Learning Environment, Instructional Delivery and Facilitation, Assessment, Professional Responsibility and Ethical Conduct, Relationships with Students, and Relationship with Parents and Community.

#### *Instructional Design and Lesson Planning*

Instructional Design and Lesson Planning contains four subscales including “Requires students to understand and demonstrate skills and competencies” (Brevard Public Schools, 2011, Appendix 35). Each subscale is scored on a 4-point Likert scale ranging from unsatisfactory to distinguished by both the teacher and the administrator. The teacher’s scores are averaged together, and the administrator’s scores are averaged together to arrive at two separate overall scores for this performance area. Space is also

provided for both the teacher and administrator to provide feedback in the form of a comment.

### *Learning Environment*

Learning Environment contains eight subscales including “Models and teaches clear, acceptable communication skills” (Brevard Public Schools, 2011, Appendix 35). Each subscale is scored on a 4-point Likert scale ranging from unsatisfactory to distinguished by both the teacher and the administrator. The teacher’s scores are averaged together, and the administrator’s scores are averaged together to arrive at two separate overall scores for this performance area. Space is also provided for both the teacher and administrator to provide feedback in the form of a comment.

### *Instructional Delivery and Facilitation*

Instructional Delivery and Facilitation contains six subscales including “Applies varied instructional strategies and resources” (Brevard Public Schools, 2011, Appendix 36). Each subscale is scored on a 4-point Likert scale ranging from unsatisfactory to distinguished by both the teacher and the administrator. The teacher’s scores are averaged together, and the administrator’s scores are averaged together to arrive at two separate overall scores for this performance area. Space is also provided for both the teacher and administrator to provide feedback in the form of a comment.

### *Assessment*

Assessment contains four subscales including “Designs and uses formative and summative assessments that lead to mastery” (Brevard Public Schools, 2011, Appendix 36). Each subscale is scored on a 4-point Likert scale ranging from unsatisfactory to

distinguished by both the teacher and the administrator. The teacher's scores are averaged together, and the administrator's scores are averaged together to arrive at two separate overall scores for this performance area. Space is also provided for both the teacher and administrator to provide feedback in the form of a comment.

### *Professional Responsibility and Ethical Conduct*

Professional Responsibility and Ethical Conduct contains five subscales including "Demonstrates professionalism" (Brevard Public Schools, 2011, Appendix 36). Each subscale is scored on a 4-point Likert scale ranging from unsatisfactory to distinguished by both the teacher and the administrator. The teacher's scores are averaged together, and the administrator's scores are averaged together to arrive at two separate overall scores for this performance area. Space is also provided for both the teacher and administrator to provide feedback in the form of a comment.

### *Relationships with Students*

Relationships with Students contains three subscales including "Displays knowledge and understanding of how students learn and applies knowledge to building positive relationships with students" (Brevard Public Schools, 2011, Appendix 37). Each subscale is scored on a 4-point Likert scale ranging from unsatisfactory to distinguished by both the teacher and the administrator. The teacher's scores are averaged together, and the administrator's scores are averaged together to arrive at two separate overall scores for this performance area. Space is also provided for both the teacher and administrator to provide feedback in the form of a comment.

### *Relationships with Parents and Community*

Relationships with Parents and Community contains ten subscales including “Promotes parent understanding of academic standards and expectations” (Brevard Public Schools, 2011, Appendix 37). Each subscale is scored on a 4-point Likert scale ranging from unsatisfactory to distinguished by both the teacher and the administrator. The teacher’s scores are averaged together, and the administrator’s scores are averaged together to arrive at two separate overall scores for this performance area. Space is also provided for both the teacher and administrator to provide feedback in the form of a comment.

The average administrator scores for each of the seven performance areas are totaled to arrive at the Total Observations Points for this instrument. A place is available for both administrator and teacher comments relative to the overall score, and signature lines are provided for both the administrator and teacher to verify that the evaluation has been discussed.

### *Professional Growth Plan Template*

Every teacher is required to write a professional growth plan (PGP). The PGP process is based on the continuous improvement cycle as shown in Figure 4 in which the administrator continuously monitors and provides feedback throughout the implementation of the plan.

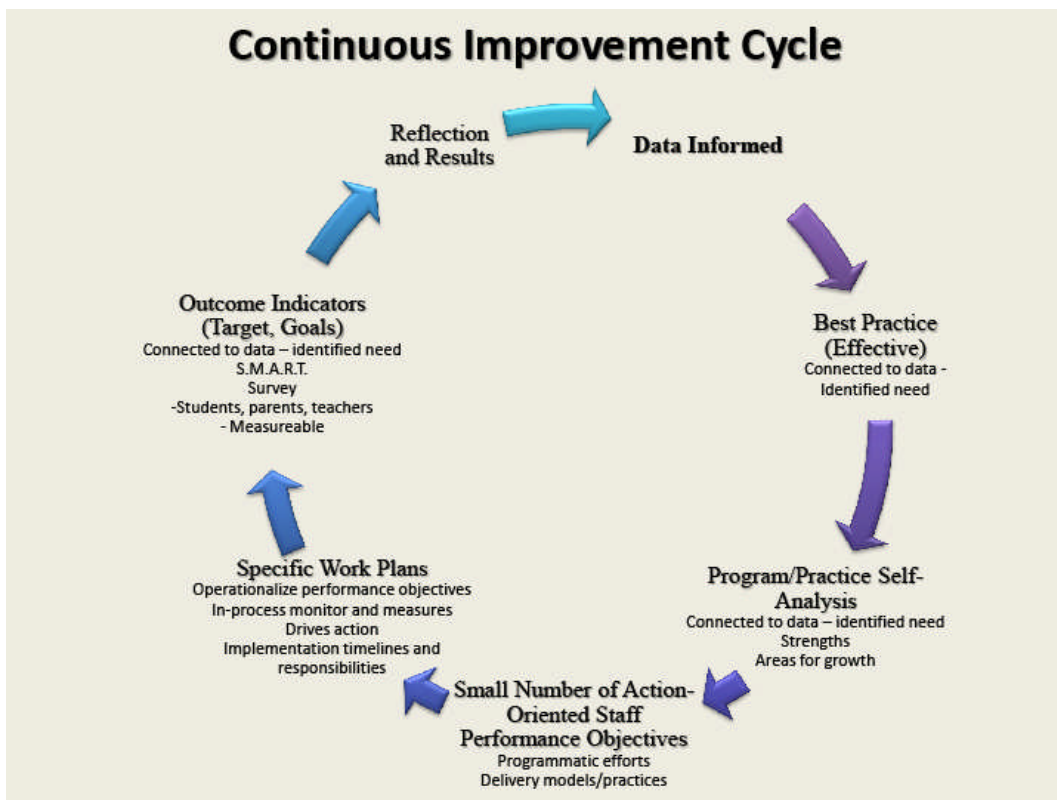


Figure 4: Continuous Improvement Cycle (Brevard Public Schools, 201, Appendix 36).

Measurable student objectives are based on the needs of the students as gathered through both quantitative and qualitative data sources. Based on a reflection of their present practices along with research and a review of school improvement goals, the teacher identifies work plan strategies, timelines, and in-process measures necessary to meet professional practice outcomes that they have established.

A template is provided for each teacher to submit his PGP to a team of three teacher leaders and an administrator for review, scoring, and approval (see Appendix K for the Instructional Professional Growth Plan Template). The instrument consists of four sections including Rationale for PGP Goal, Professional Growth Plan Goal, Work Plan Strategies, and Outcome Measures and Reflection.

### *Rationale for PGP Goal*

This part of the PGP measures how well the teacher “uses a researched, data-informed rationale to develop and analyze goals” (Brevard Public Schools, 2011, Appendix 19). The teacher provides a narrative of the data sources used to identify a need, the analysis of data, a reflection of present practice compared to research-based best practices, and how the goal is a “stretch” based on present practice.

### *Professional Growth Plan Goal*

This part of the PGP measures how well the teacher “uses a researched, data-informed rationale to develop and analyze goals” (Brevard Public Schools, 2011, Appendix 19). The teacher provides a statement regarding the new practice that will be implemented as a result of the rationale provided.

### *Work Plan Strategies*

This part of the PGP measures how well the teacher “defines qualitative and quantitative in-process measuring elements related to refined instructional practice and enhanced student mastery” (Brevard Public Schools, 2011, Appendix 20). For each work plan strategy recorded in a table format, the teacher provides the grading period(s) in which it will be implemented, specific timelines, and the in-process measures used to provide feedback.

### *Outcome Measures and Reflection*

This part of the PGP measures the impact the change in professional practice will have on student achievement (Brevard Public Schools, 2011, Appendix 20). Target goals

are identified based on changes in professional practice, and student outcomes are also identified that are specific, measureable, attainable, results-oriented, and time bound.

### *Individual Professional Growth Plan Development Rubrics*

Rubrics were written to guide the development and scoring of the PGP (see Appendix L for Individual Professional Growth Plan Development Rubrics). The rubrics provide a way of measuring teacher competency relative to the development of a professional growth plan as defined by three elements: Development of the Professional Growth Plan Goal, Work Plan Strategies, and Outcome Measures and Reflections.

#### *Development of the Professional Growth Plan Goal*

A 4-point Likert scale ranging from unsatisfactory to distinguished is provided to score this element. Descriptions of specific teacher behaviors that must be evident in order to obtain each rating are also provided. For example, in order to receive the rating of proficient, a teacher must “Show evidence of means to inform and involve students in data analysis or instructional delivery improvement efforts” and “Analyze student assessment data to develop goal/goals that are linked to classroom practice and the school improvement plan” in addition to five other descriptors (Brevard Public Schools, 2011, Appendix 19).

#### *Work Plan Strategies*

A 4-point Likert scale ranging from unsatisfactory to distinguished is provided to score this element. Descriptions of specific teacher behaviors that must be evident in order to obtain each rating are also provided. For example, in order to receive the rating of proficient, a teacher must demonstrate that “Strategies are action oriented and

sustainable” and “There is evidence of defined learning strategies and professional development to influence changes in the teacher’s instructional practices” in addition to four other descriptors (Brevard Public Schools, 2011, Appendix 20).

#### *Outcome Measures and Reflections*

A 4-point Likert scale ranging from unsatisfactory to distinguished is provided to score this element. Descriptions of specific teacher behaviors that must be evident in order to obtain each rating are also provided. For example, in order to receive the rating of proficient, a teacher must demonstrate that “The PGP target goals are connected to quantitative and/or qualitative measurable data and can be explained by changes in professional practice” and “Student outcomes and improved practice are identified” (Brevard Public Schools, 2011, Appendix 20).

#### *Professional Growth Plan Development Scoring and Feedback Form*

Individual Professional Growth Plans are scored on their development. Three teacher leaders are selected from among the faculty at each school. The selection process is left up to the individual school administrator. Using the Individual Professional Growth Plan Development Rubrics, the three teacher leaders and an administrator individually read all of their teachers’ Instructional Professional Growth Plans, which have been de-identified by number, and score them. The Professional Growth Plan Development Scoring and Feedback Form used by the teacher leaders to score the plans contains three elements; Development of PGP Goal, Work Plan Strategies, and Outcome Measures & Reflection (see Appendix M for the Professional Growth Plan Development Scoring and Feedback Form). A 4- point Likert scale ranging from unsatisfactory to



distinguished is used for the scoring of each of the three elements. Once the ratings have been determined and are recorded on the scoring form, the scores are added together to arrive at a total score. A place to write comments for each of the three elements is also provided on the form.

*Professional Growth Plan Development Administrator Final Scoring and Feedback  
Form*

The Professional Growth Plan Development Administrator Final Scoring and Feedback Form is used by the administrator to score the teacher professional growth plans (see Appendix N for the Professional Growth Plan Development Administrator Final Scoring and Feedback Form). It is the same as the teacher leaders' form. The instrument contains three elements; Development of PGP Goal, Work Plan Strategies, and Outcome Measures & Reflection. A 4-point Likert scale ranging from unsatisfactory to distinguished is used for the scoring of each of the three elements. Once the ratings have been determined and are recorded on the scoring form, the scores are added together to arrive at a total score. A place to write comments for each of the three elements is also provided on the form.

In addition to the administrator's scoring portion of the form, at the bottom of the form is a chart for calculating a teacher's final professional growth plan development score. All teacher leader scores are taken from their Professional Growth Plan Development Scoring and Feedback forms and transcribed onto this form. A final score is calculated based on the following formula: "The three teacher-leader assessments will be added together and then averaged for a sub-final score. The administrator assessment will be added to the teacher-leader sub-final score and then divided by two. The final

PGP points will then be rounded to the nearest tenth or a decimal” (Brevard Public Schools, 2011, p. 19). This is the form that is returned to the teacher after being signed by an administrator.

#### *Mid-Year Conference Form*

A mid-year conference is held by the administrator with each teacher to discuss progress on the PGP goals. This information is recorded on a Mid-Year Conference Form (see Appendix O for the Mid-Year Conference Form). This instrument contains two questions; “How are you progressing on your PGP goal?” and “Are you meeting or not meeting the goals you established” (Brevard Public Schools, 2011, Appendix 41)? The instrument was designed to measure a teacher’s progress at the mid-year point, encourage reflections, and promote “collaborative discussions regarding effective professional development” (Brevard Public Schools, 2011, p. 8). This form is signed by both the teacher and the administrator.

#### *Individual Professional Growth Plan Implementation Rubrics*

Rubrics were written to guide the implementation and scoring of the PGP (see Appendix P for Individual Professional Growth Plan Implementation Rubrics). The rubrics provide a way of measuring teacher competency relative to the implementation of a professional growth plan as defined by three elements: Implements the PGP with fidelity and professional practice, Seeks feedback and support and shares successful practice, and In-process monitoring.

### *Implements the PGP with Fidelity and Professional Practice*

A 4-point Likert scale ranging from unsatisfactory to distinguished is provided to score this element. Descriptions of specific teacher behaviors that must be evident in order to obtain each rating are also provided. For example, in order to receive the rating of proficient, a teacher must demonstrate “Consistently participates in professional development” and “While new instructional strategies are implemented, they are inconsistently integrated into lessons” in addition to one other descriptor (Brevard Public Schools, 2011, Appendix 21).

### *Seeks Feedback and Support and Shares Successful Practice*

A 4-point Likert scale ranging from unsatisfactory to distinguished is provided to score this element. Descriptions of specific teacher behaviors that must be evident in order to obtain each rating are also provided. For example, in order to receive the rating of proficient, a teacher must “Share successful practice when asked to do so” and “Seek feedback and support” in addition to one other descriptor (Brevard Public Schools, 2011, Appendix 21).

### *In-Process Monitoring*

A 4-point Likert scale ranging from unsatisfactory to distinguished is provided to score this element. Descriptions of specific teacher behaviors that must be evident in order to obtain each rating are also provided. For example, in order to receive the rating of proficient, a teacher must demonstrate that “Reflection is sporadic” and “In-process efforts provided ongoing formative data related to student mastery and success” in addition to one other descriptor (Brevard Public Schools, 2011, Appendix 21).

### *Professional Growth Plan Implementation Scoring and Feedback Form*

The Professional Growth Plan Implementation Scoring and Feedback Form is used by both the teacher and administrator at the end of the implementation period to measure a teacher's level of implementation (see Appendix Q for the Professional Growth Plan Implementation Scoring and Feedback Form). The instrument contains three elements; Implements the PGP with fidelity and professional practice, Seeks feedback and support and shares successful practice, and In-process monitoring (Brevard Public Schools, 2011, Appendix 34). A 4-point Likert scale ranging from unsatisfactory to distinguished is used for scoring each of the three elements. Once the ratings are determined and recorded on the form, the teacher's scores are added together and the administrator's scores are added together to arrive at a total score for each of them. These two scores are then averaged together to determine a teacher's final implementation score. This form also provides space for both teacher and administrator comments as well as signature lines for both teacher and administrator.

### *Collaboration & Mutual Accountability Team Scoring Form*

It is the responsibility of the administrators at every school in Brevard County to provide their teachers with opportunities to collaborate in teams with the purpose of closing the achievement gap of the lowest 25% students in reading and/or mathematics by sharing instructional strategies and practices that align with school improvement (Brevard Schools, 2011, Appendix 22 and 49). In order to measure a teacher's collaborative effort in promoting student learning, a Collaboration & Mutual Accountability Team Scoring Form is completed by every teacher (see Appendix R for the Collaboration & Mutual Accountability Team Scoring Form).

The top part of the Collaboration & Mutual Accountability Team Scoring Form contains demographic information including school name, school/district number, team member name, team number, and date. A 5-point Likert scale ranging from 0 to 8 points is provided along with descriptions of teacher behaviors associated with each. For example, to earn 8 points on the Likert scale for collaboration and mutual accountability a teacher must “Interact with colleagues in a positive manner to promote student learning and school-wide success through team efforts, vertical and/or horizontal articulation and common assessments” along with demonstrating three additional behaviors as stated on the rubric (Brevard Public Schools, 2011). The members of each person’s team are listed at the bottom of their form. Beside each team member’s name is a place to provide their score along with a space for an optional comment. Once these forms are completed, each team member’s scores are averaged together to arrive at individual team member scores.

#### *Value-Added Growth Measure*

“Brevard Public Schools will utilize the state-adopted teacher-level student growth measure” (Brevard Public Schools, 2011, p. 19). In response to Florida Statute 1008.22 (8), the Florida Department of Education established the Student Growth Implementation Committee (SGIC) for the purpose of identifying a value-added model to be used to measure student learning growth (Florida Department of Education, 2012). This committee was comprised of 27 stakeholders that met for a period of three months starting in March 2011. Technical recommendations as to which value-added model to adopt were provided by the American Institute for Research (AIR) that was contracted by the SGIC. After evaluating several models, the SGIC recommended a covariate adjustment model. This model uses prior assessment scores as predictors for current

assessment outcomes while controlling for covariates, or variables that influence learning growth outside of a teacher's control. After simulated analyses using various combinations of covariates, the following covariates were selected to be included in the model:

- up to two years of prior achievement scores,
- the number of subject-relevant courses in which the student is enrolled that are linked to an FCAT test by course code,
- Students with Disabilities (SWD) status receiving special education services,
- English Language Learners (ELL) status enrolled in a program for less than two years,
- gifted status,
- attendance, or number of days present,
- mobility, or number of transitions within the school year,
- difference from modal age of students in that grade across the state,
- class size, or the number of students linked to a teacher, and
- homogeneity of students' entering test scores in the class as a continuous variable based on the interquartile range (Florida Department of Education, 2011, p. 14).

Each student's VAM score was obtained by comparing the student's actual performance to their predicted performance with the difference being attributed to the teacher. A positive score indicated that a student did better than predicted, and a negative score indicated that a student did worse than predicted. A value-added score of zero meant that a student performed exactly as predicted. A teacher's VAM score was an aggregate of the students' scores they taught. Added to this score was 50 percent of the school effect. Therefore, the following formula was used to calculate a teacher's VAM:

Teacher VAM = Teacher Effect + .50 School Effect.

The value-added component for the Brevard County Instructional Personnel Performance Appraisal Instrument is worth 35 points. In converting value-added scores to points, the school district examined the 2010-2011 VAM scores and established VAM cut scores for elementary, middle, high school, and alternative schools (see Appendix S

for Value Added Measures). J. Carr (personal communication, January 14, 2013)

explained how this was accomplished:

- Teacher VAM estimates from 2010-2011 school year were placed in a frequency distribution and analyzed for outliers at both ends of the scale.
- The district, during the evaluation development phase, identified that a score of 28 should be the 0 VAM estimate point of a frequency distribution.
- The range of each scale was established at the 28 first, making sure that 80% of the teachers in the district were at the 28 level or above.
- Once the range was established, the remaining scale scores 0-35 were established.
- This process was done for elementary, middle, high, and alternative learning sites. The frequency distribution for the middle, high, and alternative learning sites were shown to be so similar that a single scale would be used for all three school types. Elementary school frequency distributions showed to be uniquely different to warrant a different scale.
- The teacher VAM estimate for 2011-2012 were received from the state and passed through the same process as above to validate that the scale was stable. While seeing minor variations, it was identified that the scales for elementary, middle, high, and alternative learning sites were consistent, stable, and at the level identified by the district.

The data were presented for each of the four levels using charts. The charts list the VAM score ranges along with their corresponding points. For example, a teacher at the elementary level receiving a VAM score of .47 would earn 31 points on the VAM portion of their evaluation.

### Data Analysis

The problem statement can be summarized by the question “To what extent, if any, is there a relationship between the professional practices component ratings and the VAM score on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument?” The study was guided entirely by the following research questions:

1. To what extent, if any, is there a relationship between professional practices and value-added assessment scores of instructional personnel on the School

Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

2. To what extent, if any, is there a relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

3. To what extent, if any, is there a relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

4. To what extent, if any, is there a relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

5. To what extent, if any, is there a relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional development and value-added assessment scores of instructional personnel on the School Board of



Brevard County Instructional Personnel Performance Appraisal System  
Instrument at Grades 4-10.

6. To what extent, if any, is there a relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

7. To what extent, if any, is there a relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

8. To what extent, if any, is there a relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

In determining whether a relationship existed between the professional practices component ratings and the value-added measurement score as well as the direction and strength of the relationship, a Pearson Product-Moment Correlation was the statistical method available for use. According to Lomax (2007), the variables must be interval or ratio data. The variables for the study were all interval data. Lomax also pointed out that the correlation assumes that the relationship between the two variables is a linear relationship.

The correlation coefficient  $r$  indicated the direction and strength of the relationship between the two variables. The relationship was either positive or negative indicated by the coefficient's sign. The strength fell on a scale of +1.0 to -1.0 where 0 indicated a weak relationship,  $\pm .5$  a moderate relationship, and  $\pm 1.0$  a strong relationship.

### Summary

Chapter 3 detailed the methodology for the quantitative study. Included within this chapter were an overview, the research design, and a description of the population. Additionally, methods of data collection were identified, and the instrumentation used to collect the data was detailed. Finally, the statistical analysis that was used to respond to the research questions was fully explained. The study was submitted to the Institutional Review Board for approval.

## CHAPTER IV ANALYSIS OF DATA

This study evaluated the construct validity of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. In order to determine its construct validity, it was necessary to establish if there was a relationship between the professional practices components of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument and student learning. For the purpose of this study, the professional practices components included the Formative Evaluation Observation Component and the Continuous Professional Improvement Component, and student learning was determined by a teacher's aggregated VAM score. The population of Brevard County Grade 4-10 teachers who received an aggregated VAM score based on FCAT scores associated with their instruction was used for the analysis. The statistical test Pearson Product-Moment Correlation was appropriately used to address the problem of the study. This chapter provides descriptive statistics for the population followed by data analysis results for the eight research questions included in this study.

### Descriptive Statistics

#### *Population*

During the 2011-2012 school year, which was the focus of this study, there were 85 schools in Brevard County excluding charter schools, adult education centers, alternative learning centers, virtual schools, and the school board department. The

composition of these schools included 57 elementary schools, 12 middle schools, 4 junior/senior high schools, and 12 high schools. From these 85 schools, this study was delimited to all fulltime instructional personnel in Grades 4-10 in which reading and/or mathematics was taught, the FCAT was administered, and they received a teacher aggregated VAM score as opposed to a school aggregated VAM score or a combination thereof. The following criteria were required by an instructional employee in order to receive a teacher aggregated VAM score: (a) they were hired as a first-year teacher for the 2011-2012 school year, (b) they had a combination of two VAM estimate scores for any two years, or (c) they had three years of VAM estimate scores (Brevard Public Schools, 2011).

The total number of instructional personnel who met the criteria for participation was 1,138. The majority of participants were elementary school teachers ( $n = 736$ , 64.68%), followed by middle school teachers ( $n = 177$ , 15.55%), high school teachers ( $n = 166$ , 14.59%), and junior/senior high school teachers ( $n = 59$ , 5.18%). The population included mostly females ( $n = 984$ , 86.47%) as compared to males ( $n = 154$ , 13.53%). The majority of the ethnic makeup of the population was White, Non-Hispanic ( $n = 1,008$ , 88.58%), followed by Black, Non-Hispanic ( $n = 70$ , 6.15%), Hispanic ( $n = 51$ , 4.48%), Asian or Pacific Islander ( $n = 6$ , .53%), and American Indian or Alaskan ( $n = 3$ , .26%). Teaching experience among the population had a range of 0-42, with 0 being the minimum and 42 being the maximum years of experience, and a mean of 13.21. The majority of the teachers fell within the range of 0-10 years of teaching experience ( $n = 505$ , 44.38%), followed by 11-20 years of teaching experience ( $n = 413$ , 36.29%), 21-30

years of teaching experience ( $n = 159$ , 13.97%), 31-40 years of teaching experience ( $n = 59$ , 5.18%), and 41-46 years of teaching experience ( $n = 2$ , .18%).

## Findings

The problem statement used as the basis for this study was whether there was a relationship between the Evaluation Total rating and the Value-Added Assessment score on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. A set of research questions and hypotheses specifically guided this study.

### *Research Question and Hypothesis #1*

To what extent, if any, is there a relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

The VAM scores ranged from -3.71 to 1.86 with no defined potential range. The data set had a mean score of .073 ( $s = .36$ ), a median score of .036, and a variance of .130. A histogram as show in Figure 5 representing the continuum of scores, confirmed that the peak of the distribution was represented by the interval -.25 to 0. The distribution of scores was negatively skewed (-.98) with a positive kurtosis (18.39).

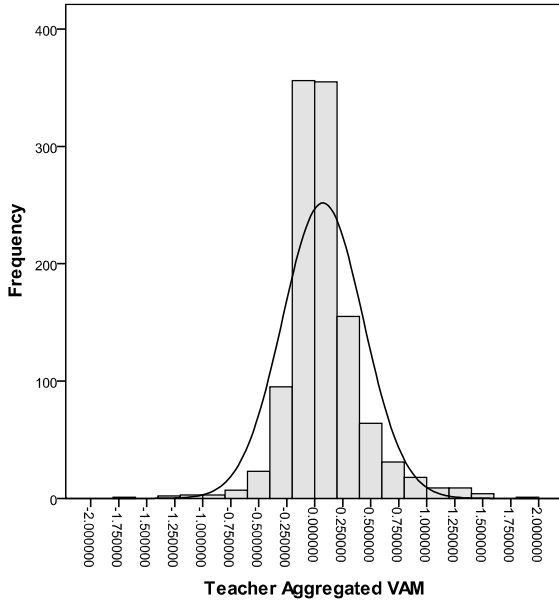


Figure 5: Teacher Aggregated VAM Histogram.

The Shapiro-Wilk test of normality was run on the VAM data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 6, the data points somewhat deviated from normality.

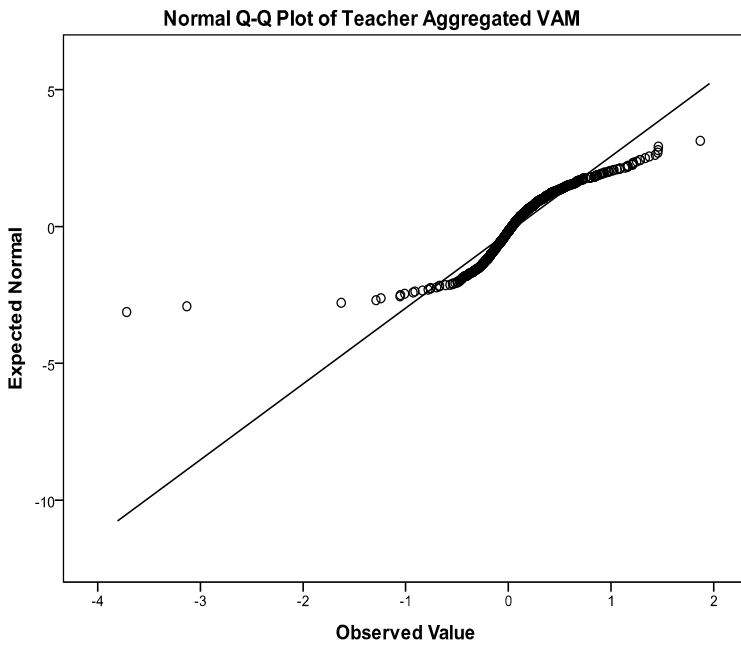


Figure 6: Normal Q-Q Plot of Teacher Aggregated VAM Scores.

In Figure 7 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower end of the distribution.

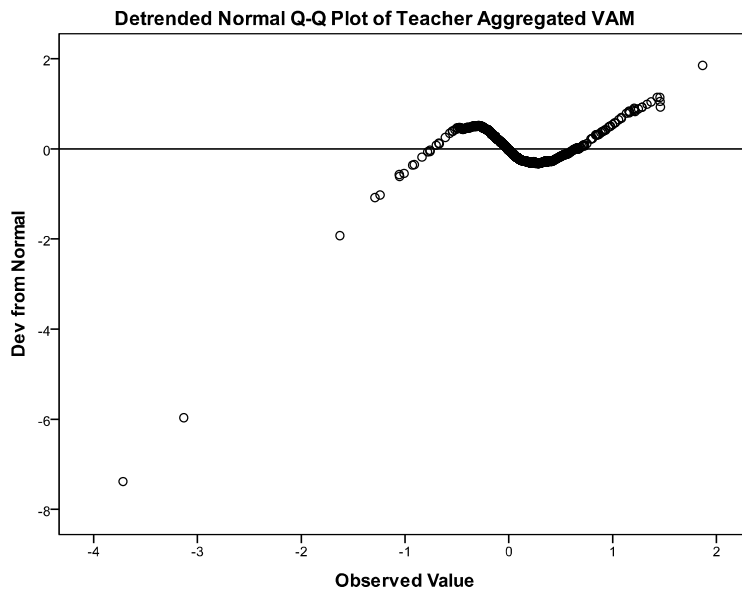


Figure 7: Detrended Normal Q-Q Plot of Teacher Aggregated VAM Scores.

A box plot, as seen in Figure 8, further indicated that the distribution was not normal, and there were quite a few outliers. Case Numbers 1,032 and 1,137 were identified as extreme outliers that were filtered out when running the correlation to determine their impact on the correlation.

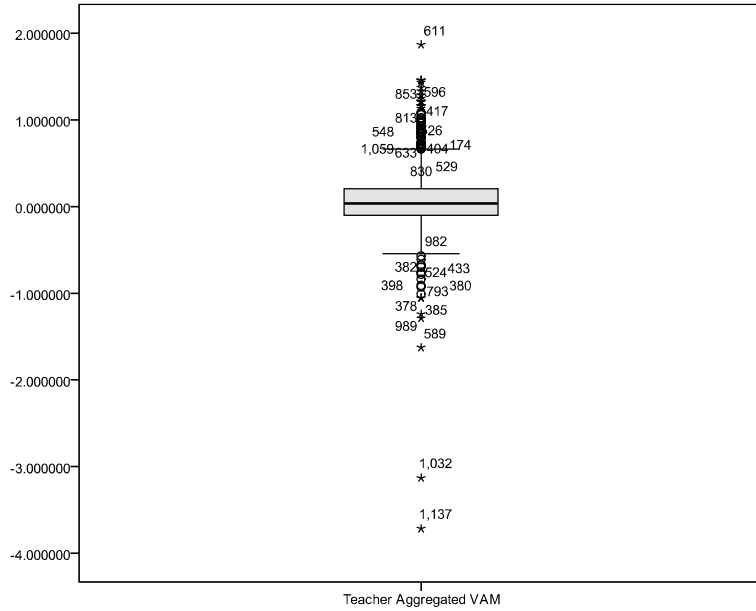


Figure 8: Box Plot of Teacher Aggregated VAM Scores.

Professional practices were measured by the Formative Evaluation Total rating on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. The ratings ranged from 23 to 47 with a potential range of 0 to 47. The data set had a mean rating of 41.59 ( $s = 3.44$ ), a median score of 42, and a variance of 11.87. A histogram, as shown in Figure 9 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 40.00 to 45.00. The distribution of ratings was negatively skewed (-.90) with a positive kurtosis (.99).



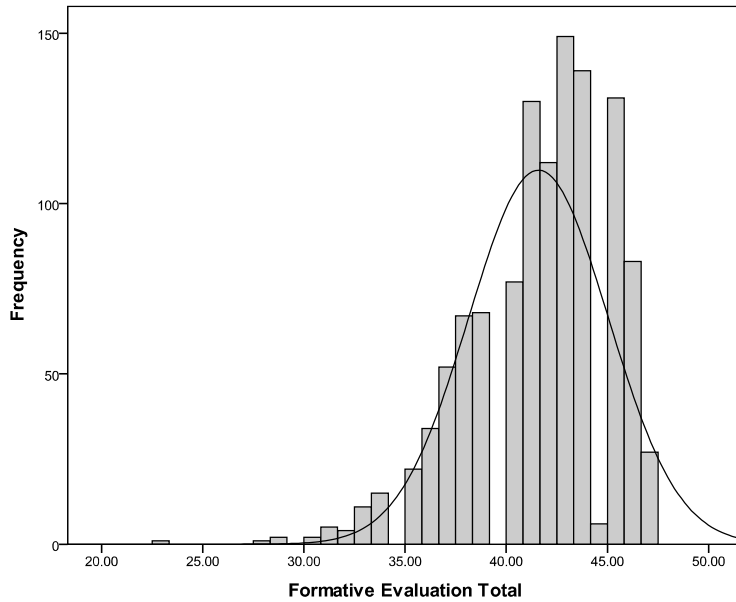


Figure 9: Professional Practices (Formative Evaluation Total) Histogram.

To determine if there was a relationship between professional practices (Formative Evaluation Total) and the VAM score of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the professional practices (Formative Evaluation Total) variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the professional practices (Formative Evaluation Total) data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 10, the data points somewhat deviated from normality.

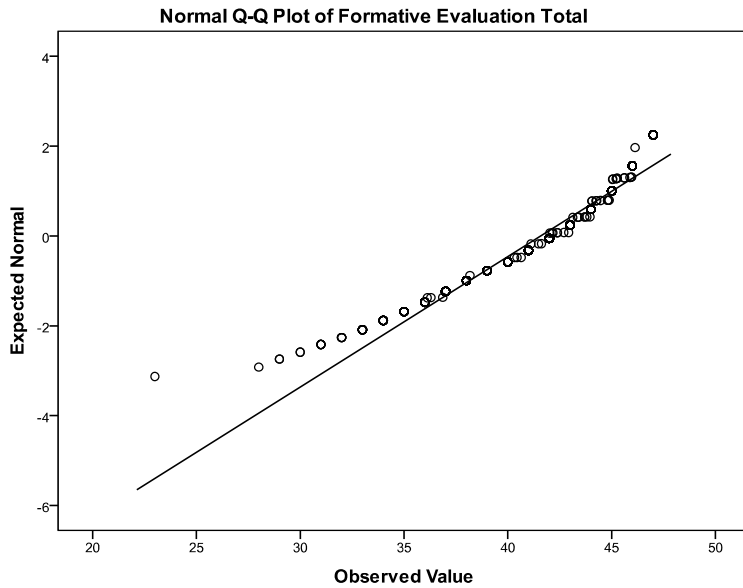


Figure 10: Normal Q-Q Plot of Professional Practices (Formative Evaluation Total) Ratings.

In Figure 11 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower end of the distribution.

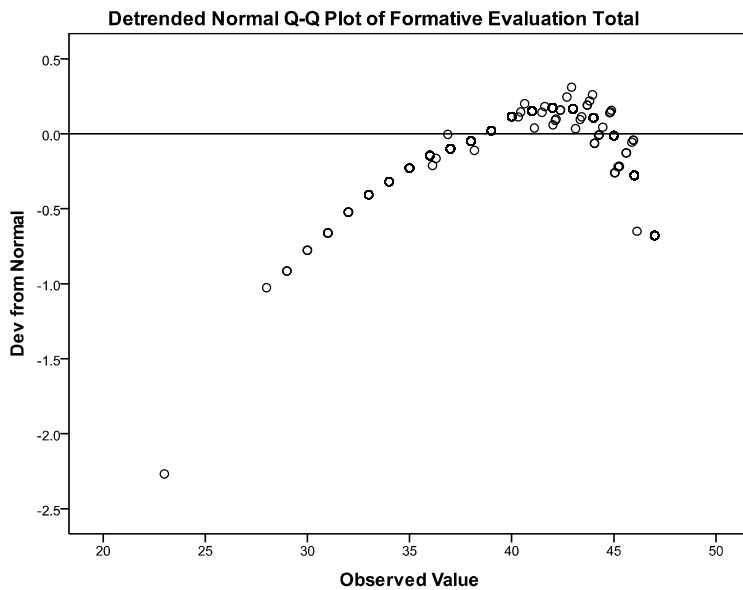


Figure 11: Detrended Q-Q Plot of Professional Practices (Formative Evaluation Total) Ratings.

A box plot, as seen in Figure 12, further indicated that the distribution was not normal, and there were quite a few outliers. Case Number 589 was identified as an

extreme outlier that was filtered out when running the correlation to determine its impact on the correlation.

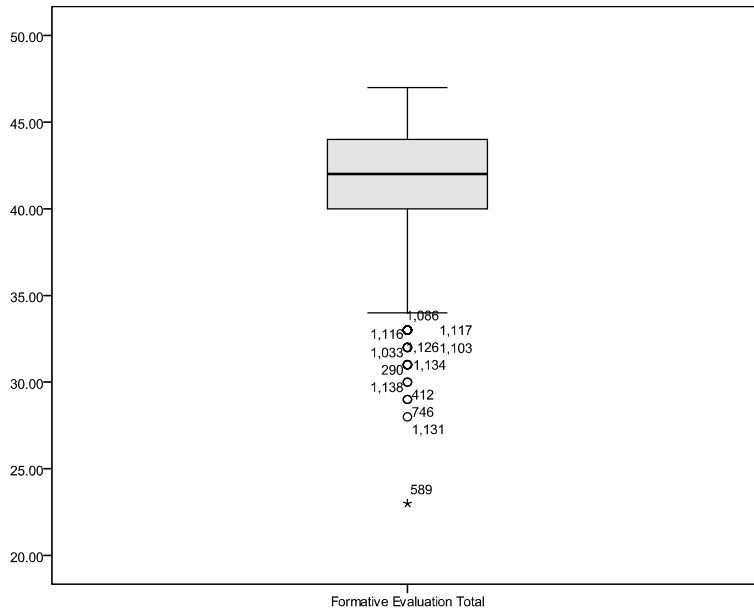


Figure 12: Box Plot of Professional Practices (Formative Evaluation Total) Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariately normally distributed or not. As can be seen in Figure 13, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

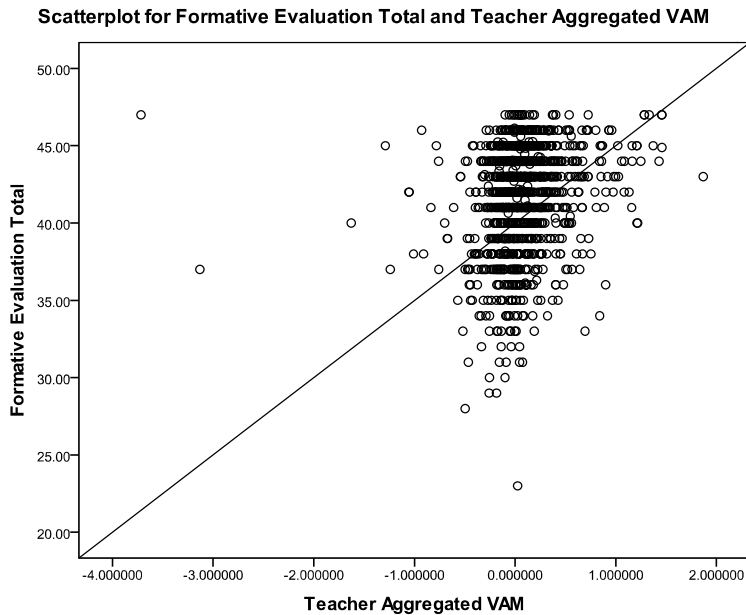


Figure 13: Scatterplot of Professional Practices (Formative Evaluation Total) Ratings and Teacher Aggregated VAM Scores.

The results of the correlational analysis indicated that there was a positive relationship between professional practices (Formative Evaluation Total) and the VAM score,  $r(1,136) = .206$ ,  $r^2 = .04$ ,  $p = .000$ . Shared variance between the two variables was approximately 4%, generally interpreted to be a small to moderate (Cohen, 1988). Only 4% of the variance in VAM scores was accounted for by the professional practices (Formative Evaluation Total) ratings. When the correlation was rerun with the three extreme outliers filtered out, there was still a small to moderate, positive correlation between the two variables,  $r(1,133) = .231$ ,  $r^2 = .05$ ,  $p = .000$ . The data analysis confirmed a significant relationship between professional practices (Formative Evaluation Total) and VAM scores; and therefore, the null hypothesis was rejected.

*Research Question and Hypothesis #2*

To what extent, if any, is there a relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

Instructional practices were measured by the Component 1 – Cumulative Total rating on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. The ratings ranged from 6.12 to 21 with a potential range of 0 to 21. The data set had a mean rating of 18.25 ( $s = 1.94$ ), a median rating of 18.65, and a variance of 3.77. A histogram, as shown in Figure 14 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 20.00 and 25.00. The distribution of ratings was negatively skewed (-.89) with a positive kurtosis (1.08).

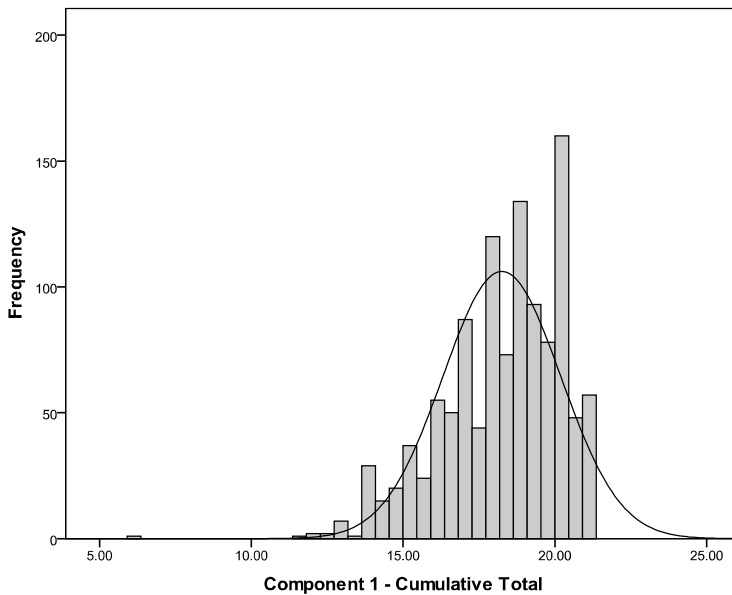


Figure 14: Instructional Practices (Component 1 - Cumulative Total) Histogram.

To determine if there was a relationship between instructional practices (Component 1 – Cumulative Total) and the VAM score of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the instructional practices (Component 1 – Cumulative Total) variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the instructional practices (Component 1 – Cumulative Total) data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 15, the data points somewhat deviated from normality at the lower end of the distribution.

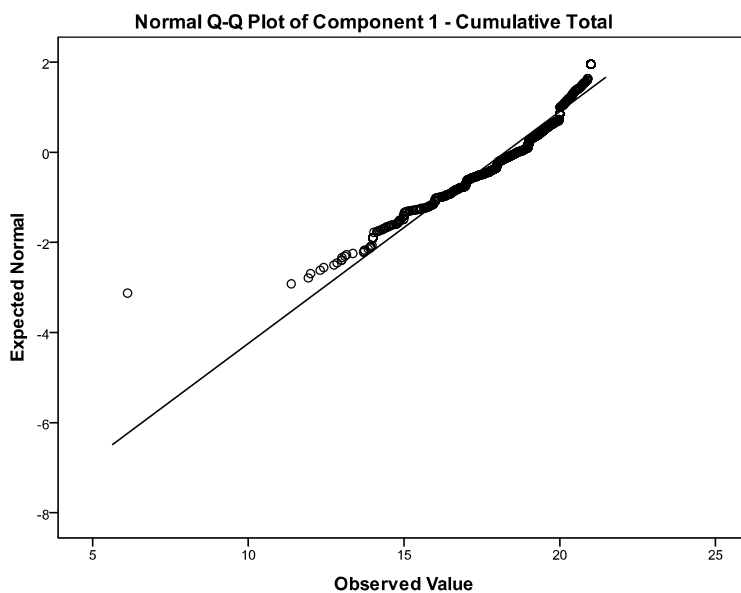


Figure 15: Normal Q-Q Plot of Instructional Practices (Component 1 - Cumulative Total) Ratings.

In Figure 16 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower end of the distribution.

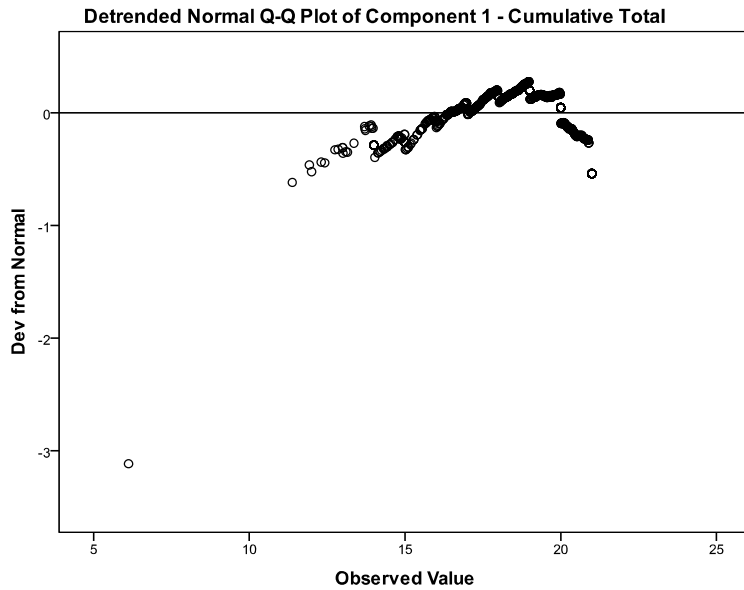


Figure 16: Detrended Normal Q-Q Plot of Instructional Practices (Component 1 - Cumulative Total) Ratings.

A box plot, as seen in Figure 17, further indicated that the distribution was not normal, and there were some outliers. Case Number 589 was identified as an extreme outlier that was filtered out when running the correlation to determine its impact on the correlation.

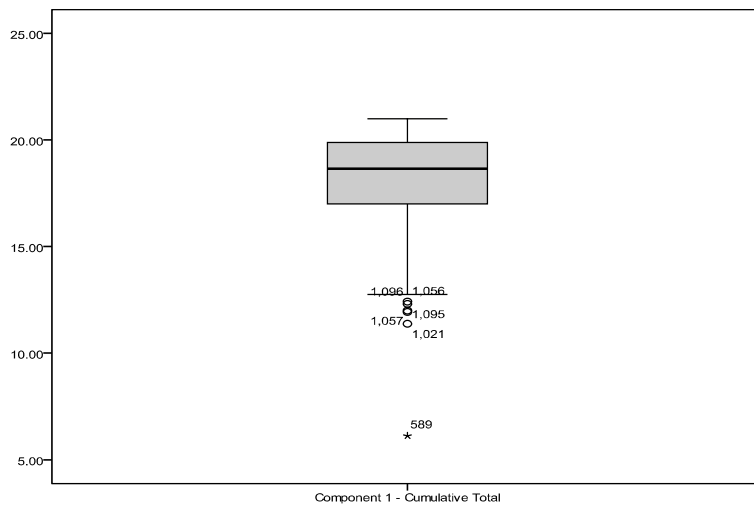


Figure 17: Box Plot of Instructional Practices (Component 1 - Cumulative Total) Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariately normally distributed or not. As can be seen in Figure 18, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

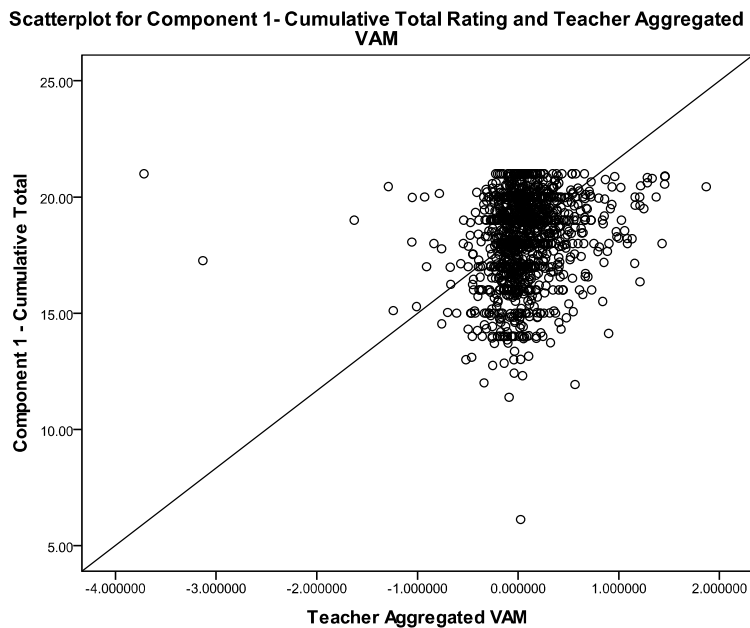


Figure 18: Scatterplot of Instructional Practices (Component 1 - Cumulative Total) Ratings and Teacher Aggregated VAM Scores.

The results of the correlational analysis indicated that there was a positive relationship between instructional practices (Component 1 – Cumulative Total) and the VAM score,  $r(1,136) = .180$ ,  $r^2 = .03$ ,  $p = .000$ . Shared variance between these two variables was approximately 3%, generally interpreted to be a small to moderate effect (Cohen, 1988). Only 3% of the variability in VAM scores was accounted for by the instructional practices (Component 1 – Cumulative Total) ratings. When the correlation



was rerun with the two extreme VAM outliers and the one extreme instructional practices outlier filtered out, there was still a small to moderate, positive correlation between the two variables,  $r(1,333) = .209$ ,  $r^2 = .04$ ,  $p = .000$ . The data analysis confirmed a significant relationship between instructional practices (Component 1 – Cumulative Total) and VAM scores; and therefore, the null hypothesis was rejected.

### *Research Question and Hypothesis #3*

To what extent, if any, is there a relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

Quality of instruction ratings ranged from 3.92 to 12 with a potential range of 0 to 12. The data set had a mean rating of 10.24 ( $s = 1.19$ ), a median rating of 10.42, and a variance of 1.42. A histogram, as shown in Figure 19 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 10.00 to 12.00. The distribution of ratings was negatively skewed (-.70) with a negative kurtosis (-.40).

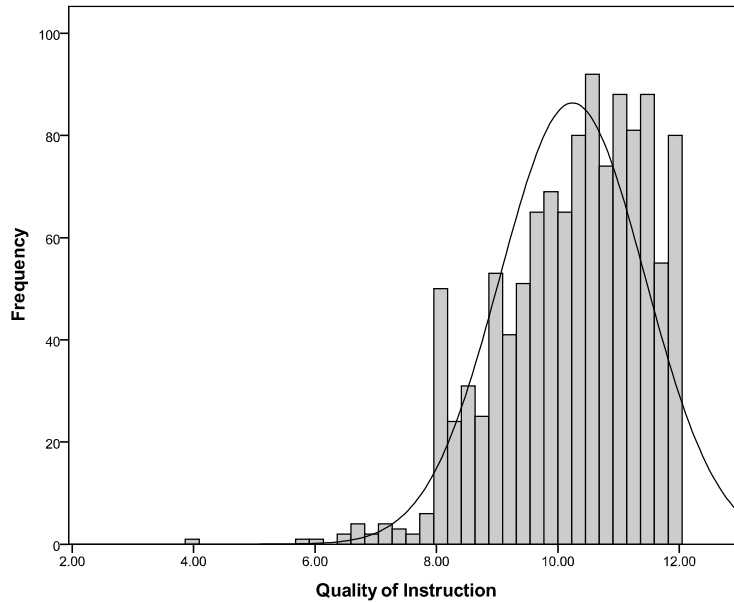


Figure 19: Quality of Instruction Histogram.

To determine if there was a relationship between the quality of instruction and the VAM score of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the quality of instruction variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the quality of instruction data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 20, the data points somewhat deviated from normality at the lower and upper ends of the distribution.

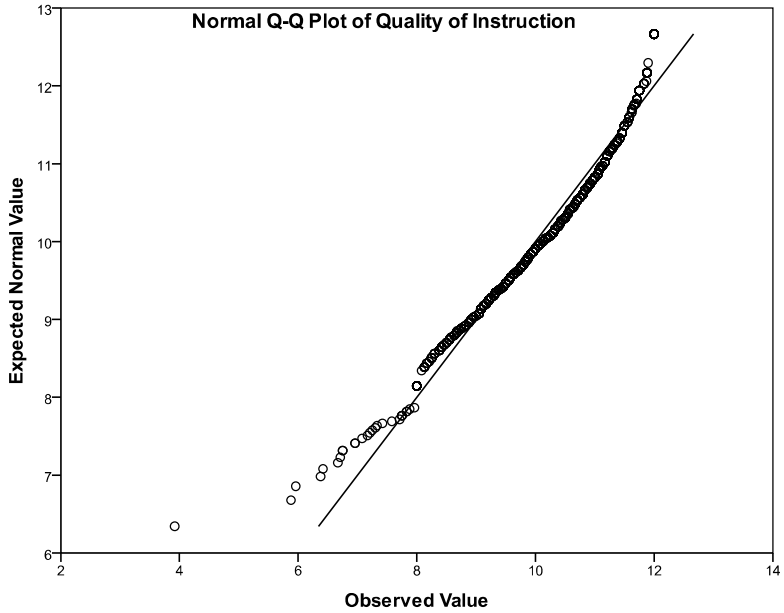


Figure 20: Normal Q-Q Plot of Quality of Instruction Ratings.

In Figure 21 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower and upper ends of the distribution.

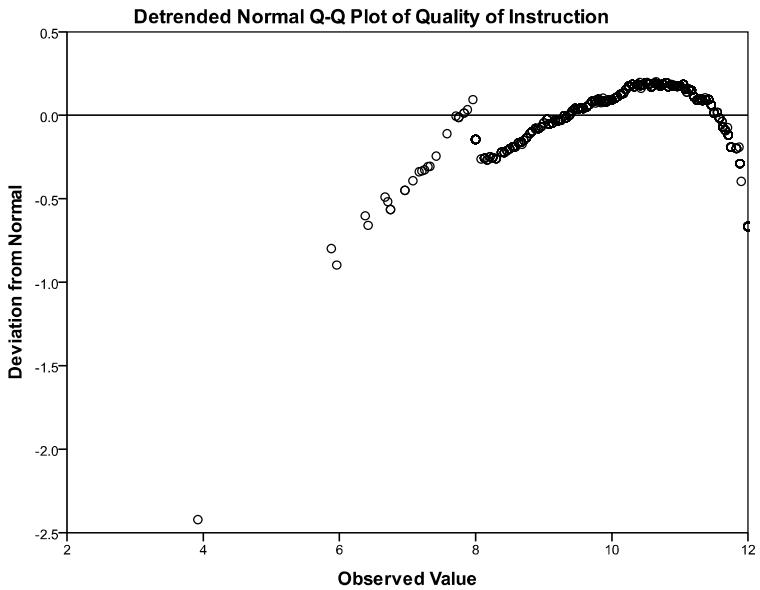


Figure 21: Detrended Normal Q-Q Plot of Quality of Instruction Ratings.

A box plot, as seen in Figure 22, further indicated that the distribution was not normal, and there were outliers. Case Number 589 was identified as an extreme outlier

that was filtered out when running the correlation to determine its impact on the correlation.

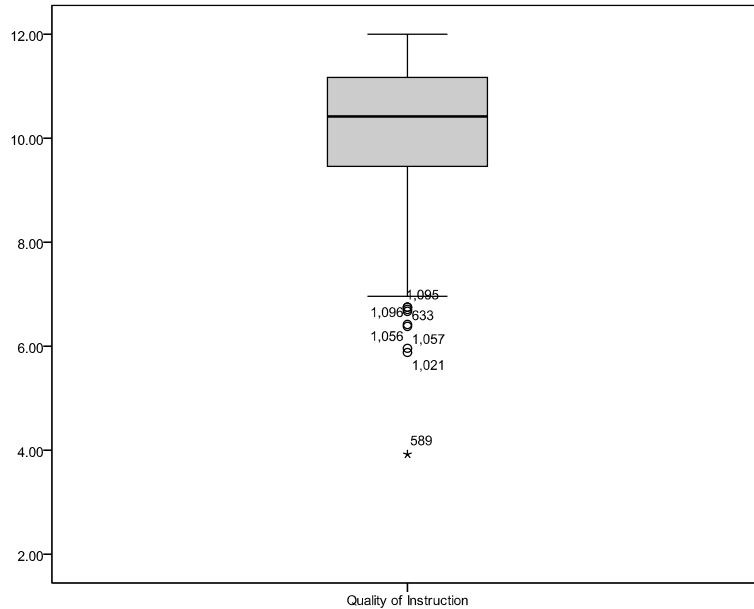


Figure 22: Box Plot of Quality of Instruction Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariately normally distributed or not. As can be seen in Figure 23, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

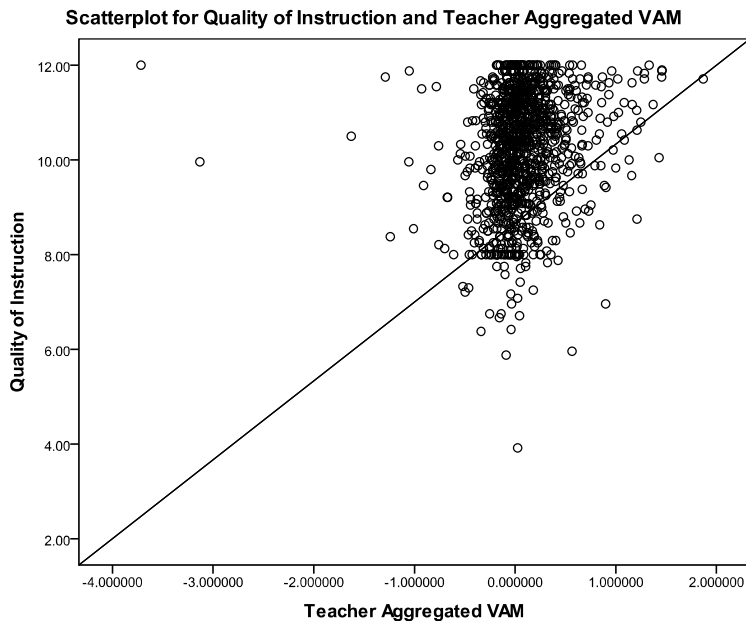


Figure 23: Scatterplot of Quality of Instruction Ratings and Teacher Aggregated VAM Scores.

The results of the correlational analysis indicated that there was a positive relationship between the quality of instruction rating and the VAM score,  $r(1,136) = .187$ ,  $r^2 = .03$ ,  $p = .000$ . Shared variance between these two variables was approximately 3%, generally interpreted to be a small to moderate effect (Cohen, 1988). Only 3% of the variability in VAM scores was accounted for by the quality of instruction ratings. When the correlation was rerun with the two extreme VAM outliers and the one extreme quality of instruction outlier filtered out, there was still a small to moderate, positive correlation between the two variables,  $r(1,133) = .218$ ,  $r^2 = .05$ ,  $p = .000$ . The data analysis confirmed a significant relationship between instructional practices and VAM scores; and therefore, the null hypothesis was rejected.

*Research Question and Hypothesis #4*

To what extent, if any, is there a relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

Professional responsibility, conduct and relationships ratings ranged from 2.20 to 9 with a potential range of 0 to 9. The data set had a mean rating of 7.96 ( $s = .83$ ), a median rating of 8.17, and a variance of .69. A histogram, as shown in Figure 24 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 8.00 to 10.00. The distribution of ratings was negatively skewed (-1.04) with a positive kurtosis (1.68).

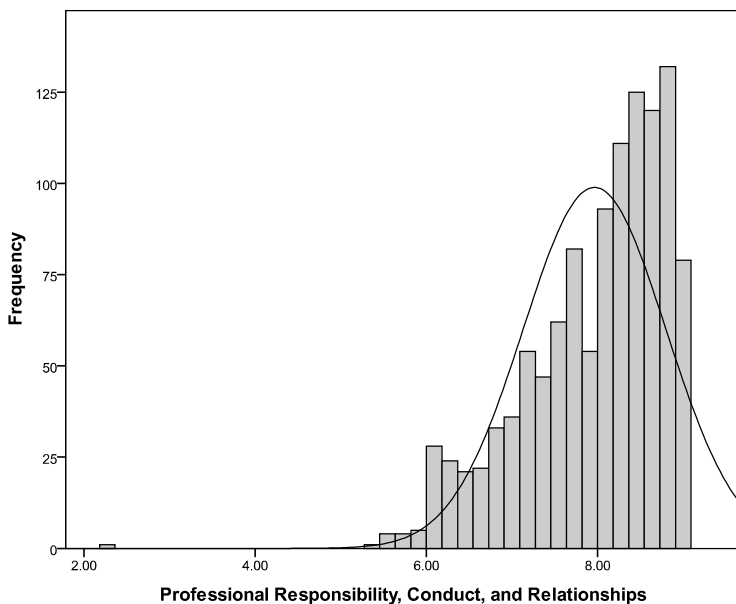


Figure 24: Professional Responsibility, Conduct and Relationships Ratings Histogram.

To determine if there was a relationship between the professional responsibility, conduct, and relationships and the VAM score of instructional personnel on the School

Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the professional responsibility, conduct, and relationships variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the Formative Evaluation Tool data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 25, the data points somewhat deviated from normality at the lower and upper ends of the distribution.

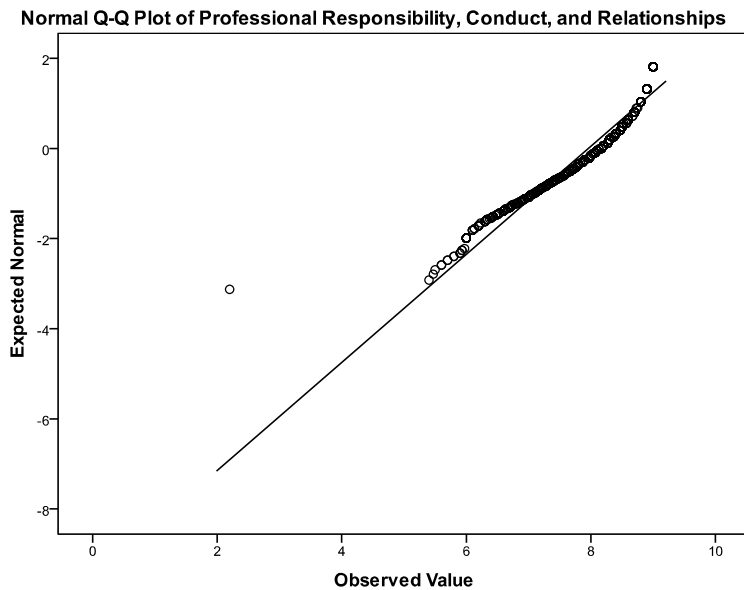


Figure 25: Normal Q-Q Plot of Professional Responsibility, Conduct and Relationships Ratings.

In Figure 26 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower and upper ends of the distribution.

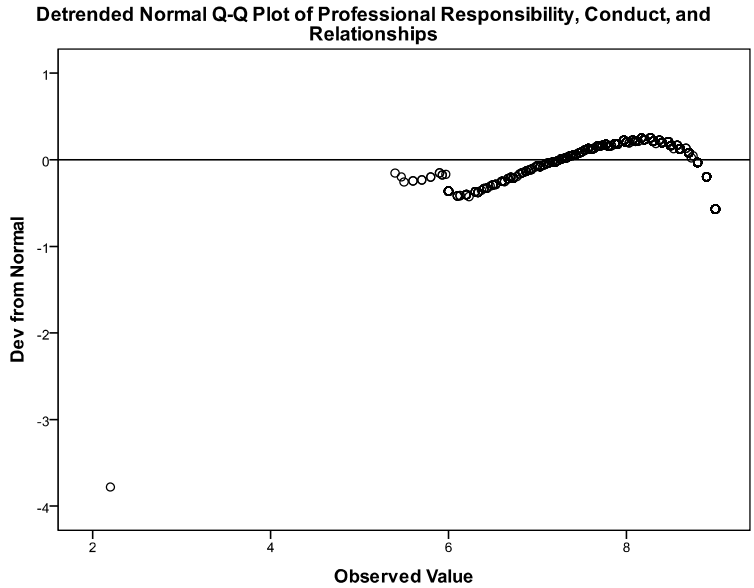


Figure 26: Detrended Normal Q-Q Plot of Professional Responsibility, Conduct and Relationships Ratings.

A box plot, as seen in Figure 27, further indicated that the distribution was not normal, and there were outliers. Case Number 589 was identified as an extreme outlier that was filtered out when running the correlation to determine its impact on the correlation.

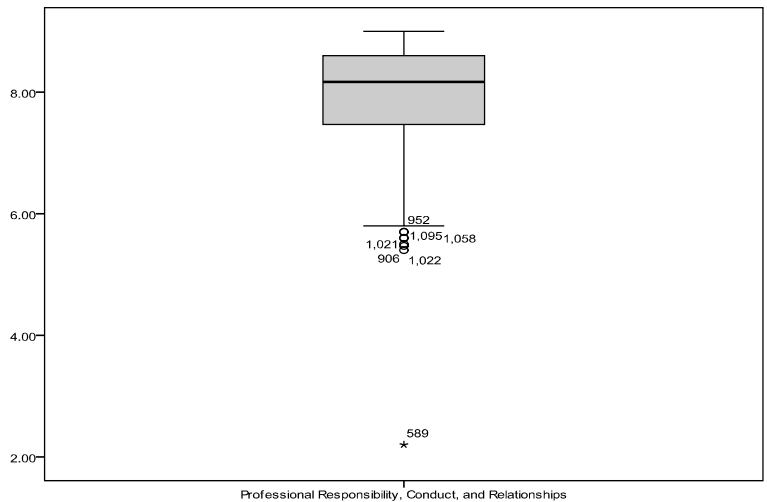


Figure 27: Box Plot of Professional Responsibility, Conduct and Relationships Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariately normally distributed or not. As



can be seen in Figure 28, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

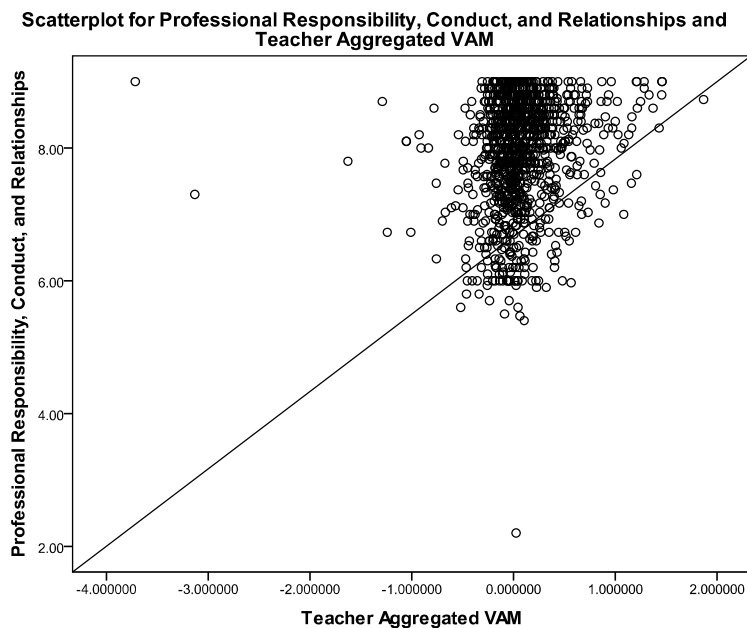


Figure 28: Scatterplot of Professional Responsibility, Conduct and Relationships Ratings and Teacher Aggregated VAM Scores.

The results of the correlational analysis indicated that there was a positive relationship between professional responsibility, conduct, and relationships and the VAM score,  $r(1,136) = .166$ ,  $r^2 = .03$ ,  $p = .000$ . Shared variance between these two variables was approximately 3%, generally interpreted to be a small to moderate effect (Cohen, 1988). Only 3% of the variability in VAM scores was accounted for by the professional responsibility, conduct, and relationships ratings. When the correlation was rerun with the two VAM extreme outliers and the one professional responsibility, conduct, and relationships extreme outlier filtered out, there was still a small to moderate, positive

correlation between the two variables,  $r(1,133) = .188$ ,  $r^2 = .04$ ,  $p = .000$ . The data analysis confirmed a significant relationship between professional responsibility, conduct, and relationships and VAM scores; and therefore, the null hypothesis was rejected.

#### *Research Question and Hypothesis #5*

To what extent, if any, is there a relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

Professional development was measured by the Component 2 – Cumulative Total rating on the School Board of Brevard County Personnel Performance Appraisal System Instrument. The ratings ranged from 8 to 26 with a potential range of 0 to 26. The data set had a mean rating of 23.37 ( $s = 2.10$ ), a median rating of 23.83, and a variance of 4.45. A histogram, as shown in Figure 29 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 25.00 to 30.00. The distribution of ratings was negatively skewed (-1.33) with a positive kurtosis (3.46).

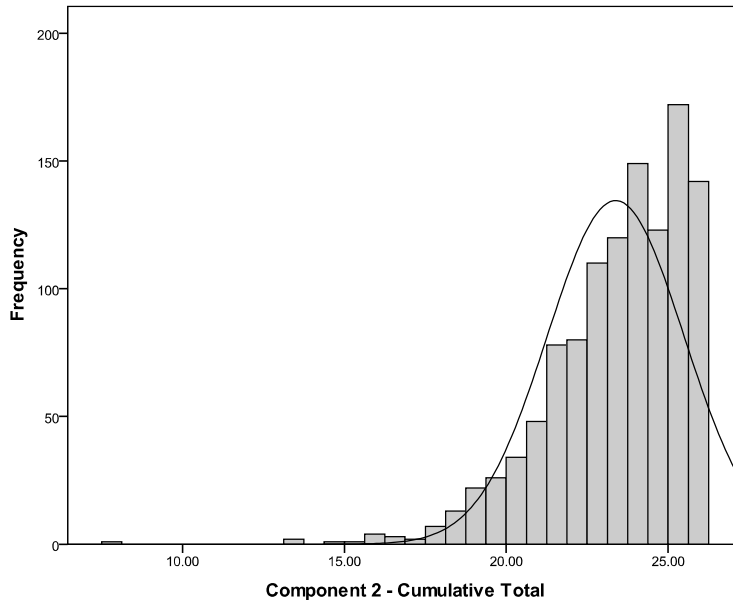


Figure 29: Professional Development (Component 2 - Cumulative Total) Ratings Histogram.

To determine if there was a relationship between professional development (Component 2 - Cumulative Total) and the VAM score of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the professional development variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the professional development (Component 2 – Cumulative Total) data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 30, the data points somewhat deviated from normality at the lower and upper ends of the distribution.

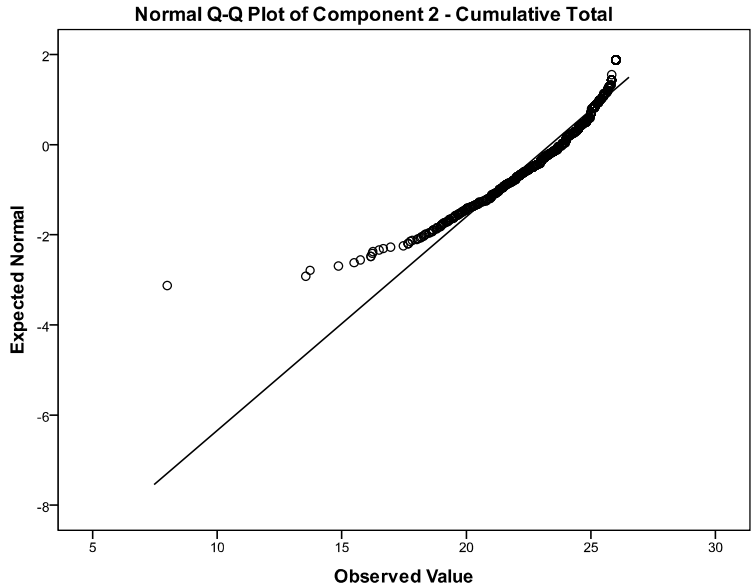


Figure 30: Normal Q-Q Plot of Professional Development (Component 2 - Cumulative Total) Ratings.

In Figure 31 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower and upper ends of the distribution.

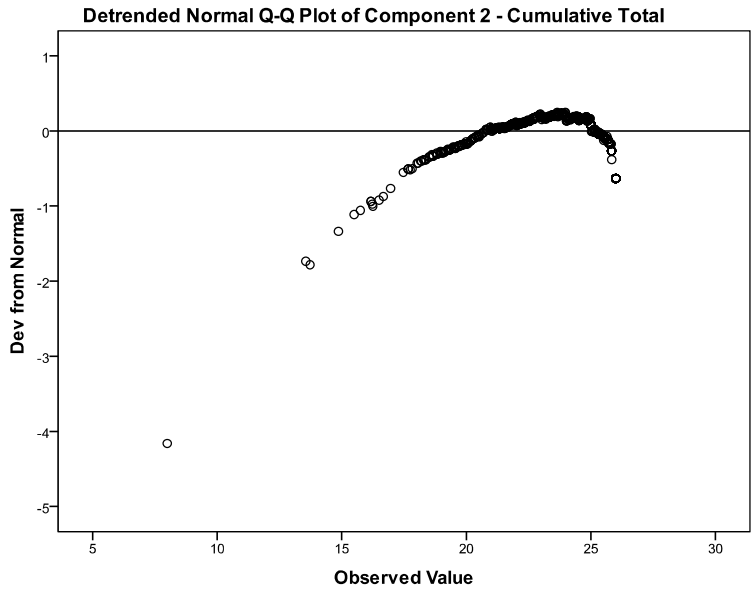


Figure 31: Detrended Normal Q-Q Plot of Professional Development (Component 2 - Cumulative Total) Ratings.

A box plot, as seen in Figure 32, further indicated that the distribution was not normal, and there were significant outliers. Case Number 1,094 was identified as an

extreme outlier that was filtered out when running the correlation to determine its impact on the correlation.

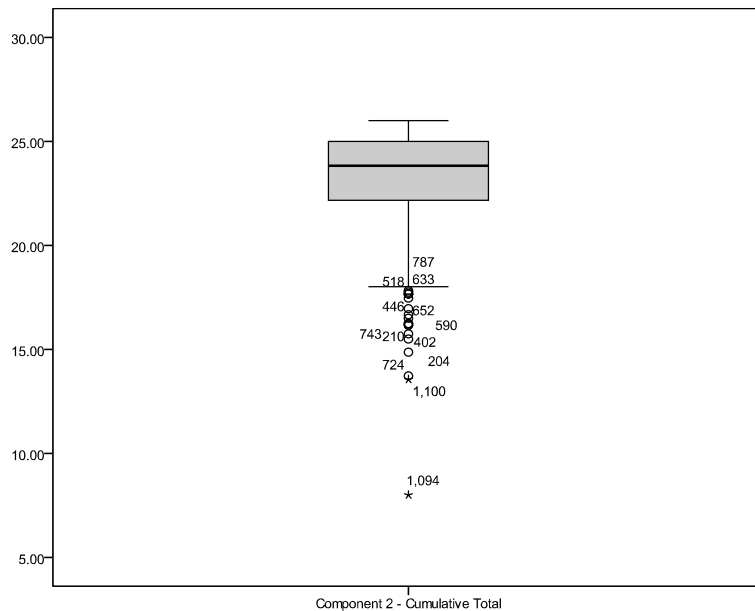


Figure 32: Box Plot of Professional Development (Component 2 - Cumulative Total) Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariate normally distributed or not. As can be seen in Figure 33, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

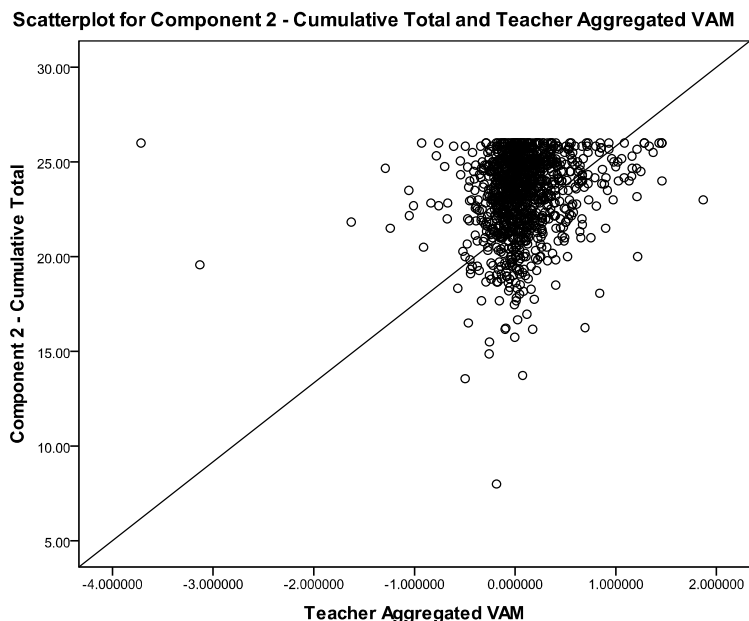


Figure 33: Scatterplot of Professional Development (Component 2 - Cumulative Total) Ratings and Teacher Aggregated VAM Scores.

The results of the correlational analysis indicated that there was a positive relationship between the professional development (Component 2 – Cumulative Total) and the VAM score,  $r(1,136) = .175$ ,  $r^2 = .03$ ,  $p = .000$ . Shared variance between these two variables was approximately 3%, generally interpreted to be a small to moderate effect (Cohen, 1988). Only 3% of the variability in VAM scores was accounted for by the professional development (Component 2 – Cumulative Total) ratings. When the correlation was rerun with the two VAM extreme outliers and the one professional development extreme outlier filtered out, there was still a small to moderate, positive correlation between the two variables,  $r(1,133) = .189$ ,  $r^2 = .04$ ,  $p = .000$ . The data analysis confirmed a significant relationship between professional development (Component 2 – Cumulative Total) and VAM scores; and therefore, the null hypothesis was rejected.

*Research Question and Hypothesis #6*

To what extent, if any, is there a relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

Action research development was measured by the Professional Growth Plan Development rating on the School Board of Brevard County Personnel Performance Appraisal System Instrument. The ratings ranged from 0 to 10 with a potential range of 0 to 10. The data set had a mean rating of 8.70 ( $s = 1.35$ ), a median rating of 9, and a variance of 1.83. A histogram, as shown in Figure 34 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 8.00 to 10.00. The distribution of ratings was negatively skewed (-1.62) with a positive kurtosis (4.12).

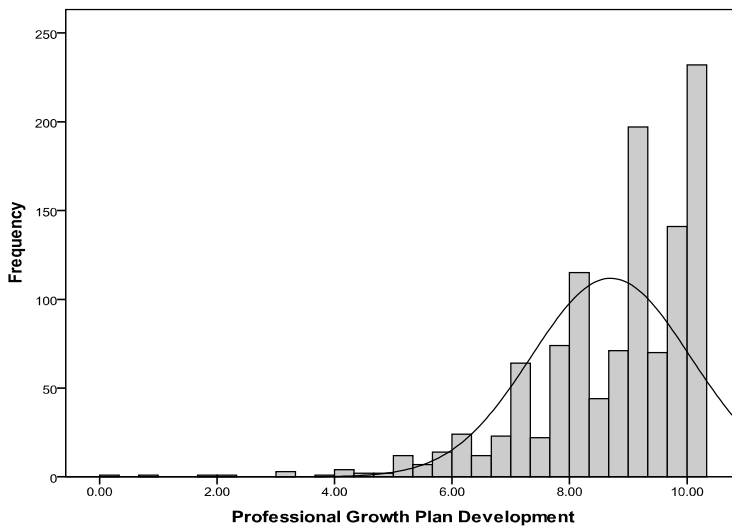


Figure 34: Action Research Development (Professional Growth Plan Development) Ratings Histogram.

To determine if there was a relationship between action research development (Professional Growth Plan Development) and the VAM score of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the action research development (Professional Growth Plan Development) variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the action research development (Professional Growth Plan Development) data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 35, the data points deviated from normality at the lower end of the distribution.

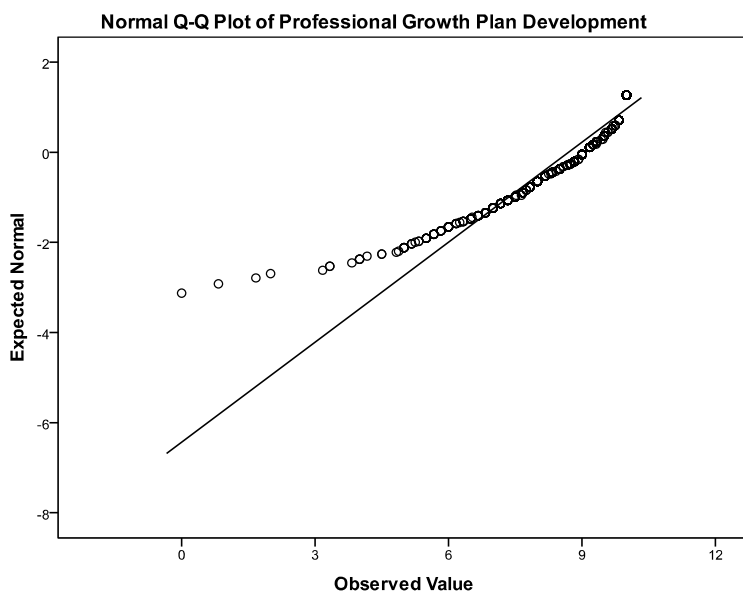


Figure 35: Normal Q-Q Plot of Action Research Development (Professional Growth Plan Development) Ratings.

In Figure 36 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower end of the distribution.



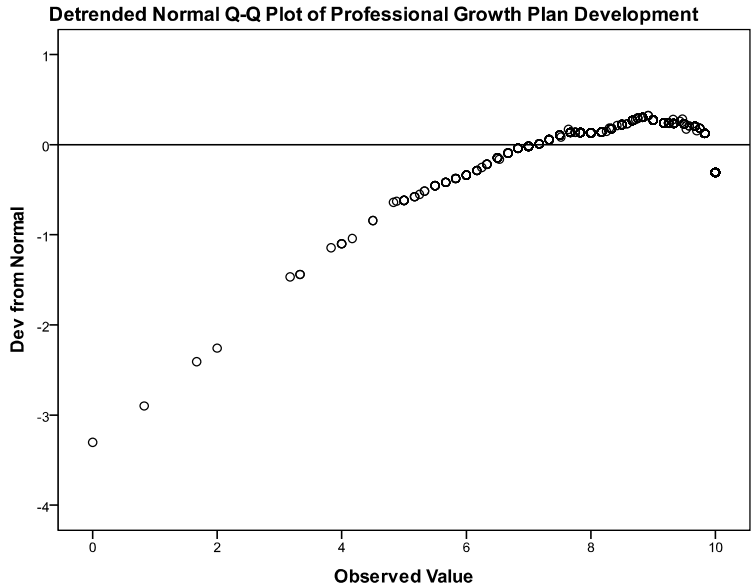


Figure 36: Detrended Normal Q-Q Plot of Action Research Development (Professional Growth Plan Development) Ratings.

A box plot, as seen in Figure 37, further indicated that the distribution was not normal, and there were significant outliers. Case Numbers 1,094, 204, 590, and 402 were identified as extreme outliers that were filtered out when running the correlation to determine their impact on the correlation.

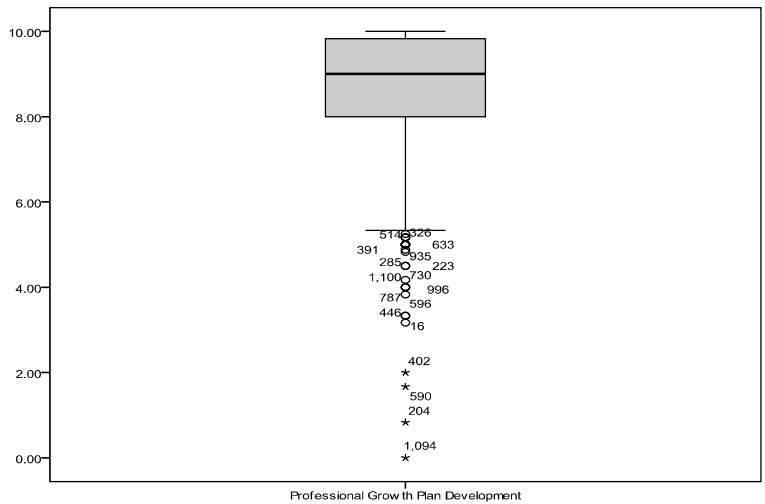


Figure 37: Box Plot of Action Research Development (Professional Growth Plan Development) Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariately normally distributed or not. As can be seen in Figure 38, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

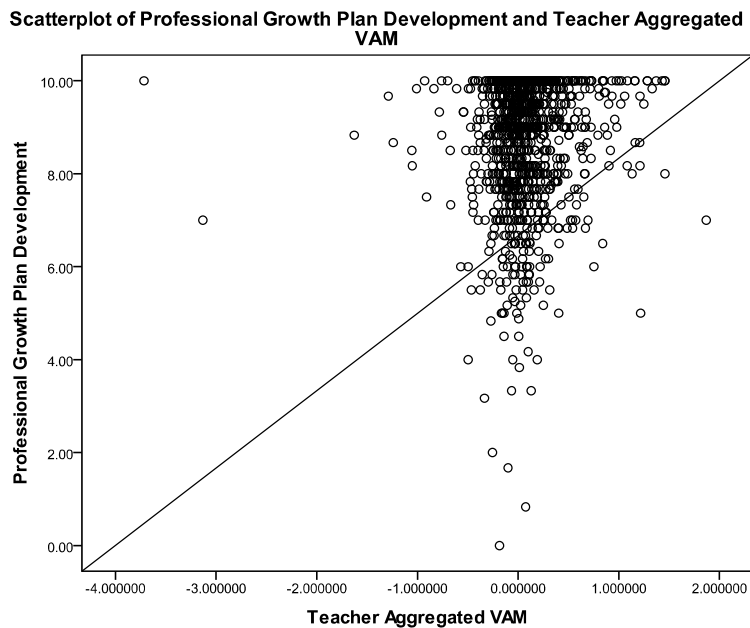


Figure 38: Scatterplot of Action Research Development (Professional Growth Plan Development) Ratings.

The results of the correlational analysis indicated that there was a positive relationship between action research development (Professional Growth Plan Development) and the VAM score,  $r(1,136) = .099$ ,  $r^2 = .01$ ,  $p = .001$ . Shared variance between these two variables was approximately 1%, generally interpreted to be a small effect (Cohen, 1988). Only 1% of the variability in VAM scores was accounted for by the action research development (Professional Growth Plan Development) ratings. When

the correlation was rerun with the two VAM extreme outliers and the four action research development extreme outliers filtered out, there was still a small, positive correlation between the two variables,  $r(1,130) = .102$ ,  $r^2 = .01$ ,  $p = .001$ . The data analysis confirmed a significant relationship between action research development (Professional Growth Plan Development) and VAM scores; and therefore, the null hypothesis was rejected.

### *Research Question and Hypothesis #7*

To what extent, if any, is there a relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

Implementation of action research was measured by the Professional Growth Plan Implementation rating on the School Board of Brevard County Personnel Performance Appraisal System Instrument. The ratings ranged from 0 to 8, with a potential range of 0 to 8. The data set had a mean rating of 6.91 ( $s = 1.19$ ), a median rating of 7, and a variance of 1.43. A histogram, as shown in Figure 39 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 6.00 to 8.00. The distribution of ratings was negatively skewed (-1.51) with a positive kurtosis (4.49).

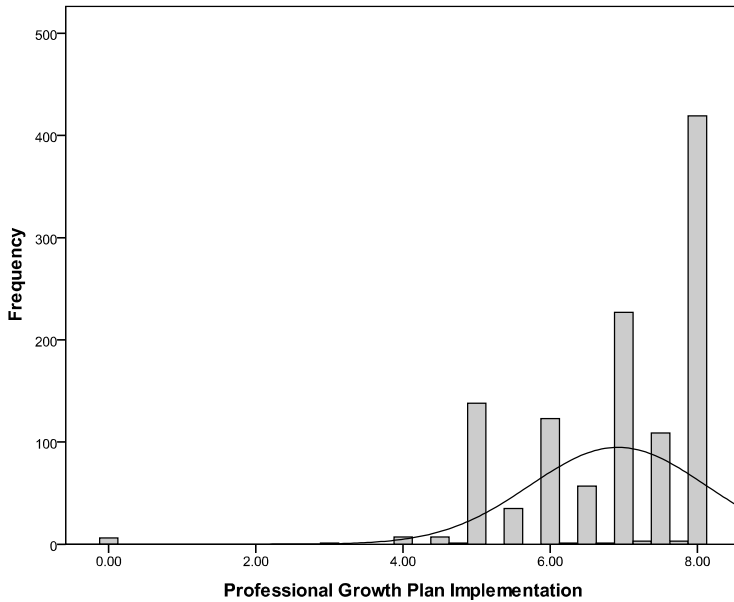


Figure 39: Implementation of Action Research (Professional Growth Plan Implementation) Ratings Histogram.

To determine if there was a relationship between the implementation of action research (Professional Growth Plan Implementation) and the VAM score of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the implementation of action research variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the Professional Growth Plan Implementation data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 40, the data points deviated from normality at the lower end of the distribution.

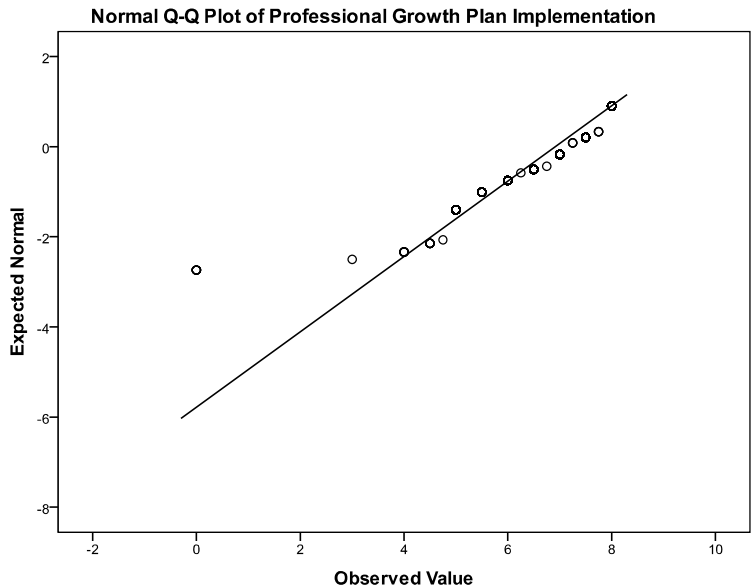


Figure 40: Normal Q-Q Plot Implementation of Action Research (Professional Growth Plan Implementation) Ratings.

In Figure 41 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower end of the distribution.

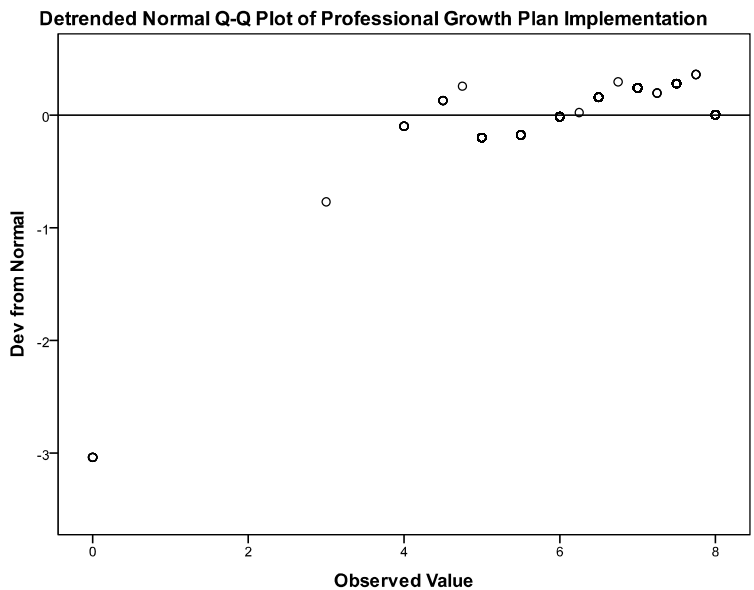


Figure 41: Detrended Normal Q-Q Plot of Implementation of Action Research (Professional Growth Plan Implementation) Ratings.

A box plot, as seen in Figure 42, indicated that the distribution was normal with only three outliers. Case Numbers 652, 1,094, and 518 were identified as outliers that

were filtered out when running the correlation to determine their impact on the correlation.

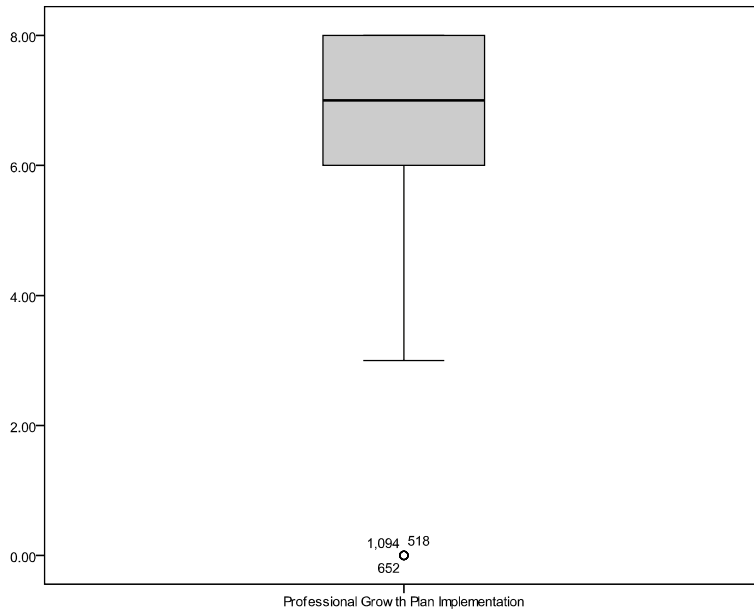


Figure 42: Box Plot of Implementation of Action Research (Professional Growth Plan Implementation) Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariate normally distributed or not. As can be seen in Figure 43, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

The results of the correlational analysis indicated that there was a positive relationship between the implementation of action research (Professional Growth Plan Implementation) and the VAM score,  $r(1,136) = .161$ ,  $r^2 = .03$ ,  $p = .000$ . Shared variance

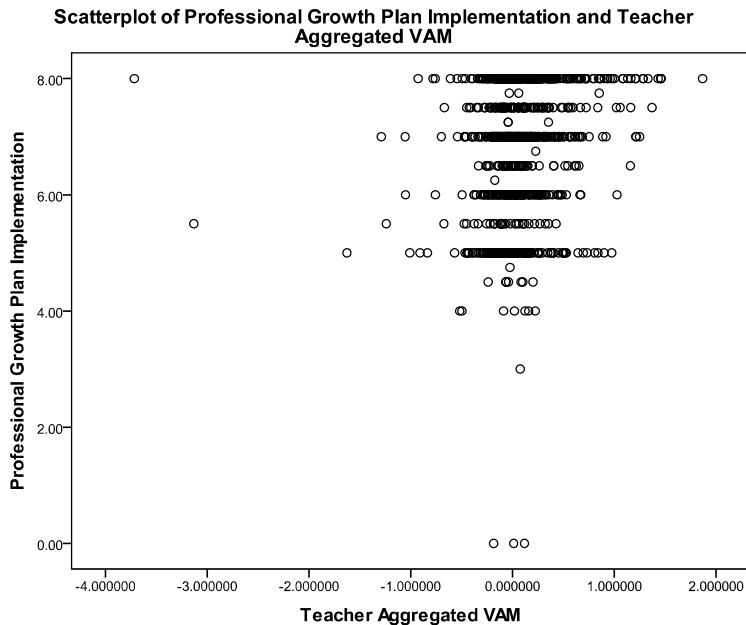


Figure 43: Scatterplot of Implementation of Action Research (Professional Growth Plan Implementation) Ratings.

between these two variables was approximately 3%, generally interpreted to be a small to moderate effect (Cohen, 1988). Only 3% of the variability in VAM scores was accounted for by the implementation of action research (Professional Growth Plan Implementation) ratings. When the correlation was rerun with the two VAM extreme outliers and the three implementation of action research outliers filtered out, there was still a small to moderate, positive correlation between the two variables,  $r(1,131) = .175$ ,  $r^2 = .03$ ,  $p = .000$ . The data analysis confirmed a significant relationship between implementation of action research (Professional Growth Plan Implementation) and VAM scores; and therefore, the null hypothesis was rejected.

#### *Research Question and Hypothesis #8*

To what extent, if any, is there a relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of

## Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

Collaborative inquiry was measured by the Collaboration and Mutual Accountability rating on the School Board of Brevard County Personnel Performance Appraisal Instrument. The ratings ranged from 3.2 to 8 with a potential range of 0 to 8. The data set had a mean rating of 7.76 ( $s = .51$ ), a median rating of 8, and a variance of .26. A histogram, as shown in Figure 44 representing the continuum of ratings, confirmed that the peak of the distribution was represented by the interval 7.0 to 8.0. The distribution of ratings was negatively skewed (-3.37) with a positive kurtosis (15.76).

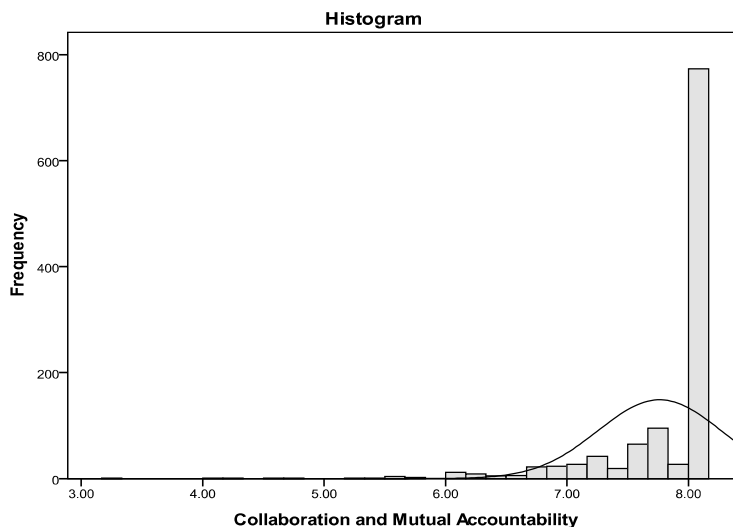


Figure 44: Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings Histogram.

To determine if there was a relationship between the collaborative inquiry (Collaboration and Mutual Accountability) and the VAM score of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument, a correlation coefficient was calculated. First, the normality of the



collaborative inquiry (Collaboration and Mutual Accountability) variable was explored to provide a clear understanding of its distribution shape and identify any outliers that needed to be eliminated. The Shapiro-Wilk test of normality was run on the collaborative inquiry (Collaboration and Mutual Accountability) data resulting in a significant level of .000 indicating a lack of normality. A Q-Q Plot was also run to determine the normality of the data. As seen in Figure 45, the data points deviated from normality especially at the lower end of the distribution.

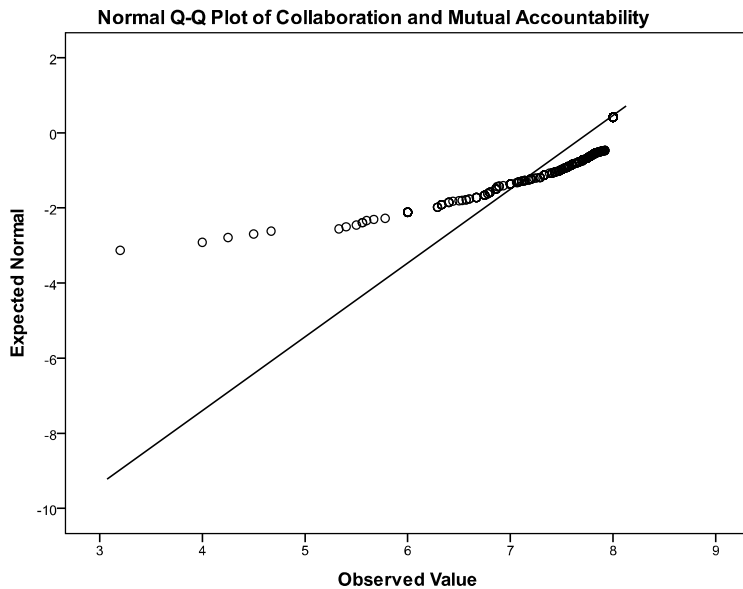


Figure 45: Normal Q-Q Plot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings.

In Figure 46 the Detrended Normal Q-Q Plot also indicated a deviation from normality especially at the lower and upper ends of the distribution.

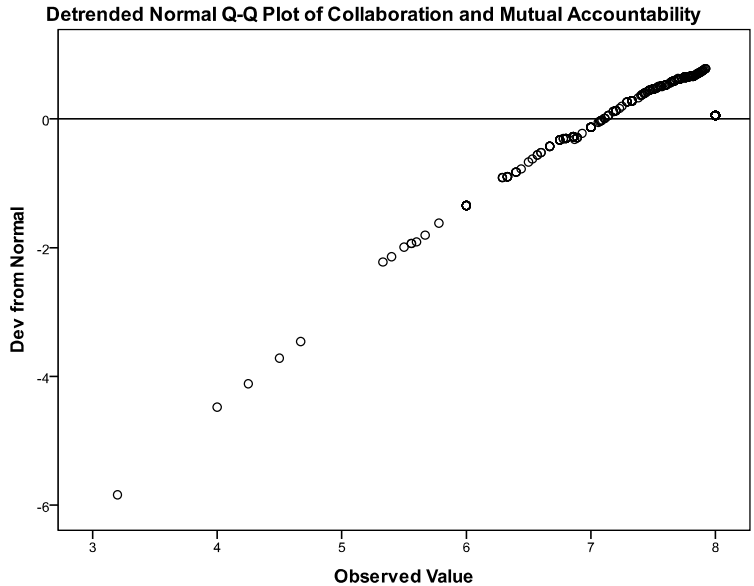


Figure 46: Detrended Normal Q-Q Plot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings.

A box plot, as seen in Figure 47, indicated that the distribution was not normal with significant outliers. Since there were so many extreme outliers, Case Numbers 1,134, 1,135, 1,136, 1,137, and 1,138 being the most extreme outliers were filtered out when running the correlation to determine their impact on the correlation.

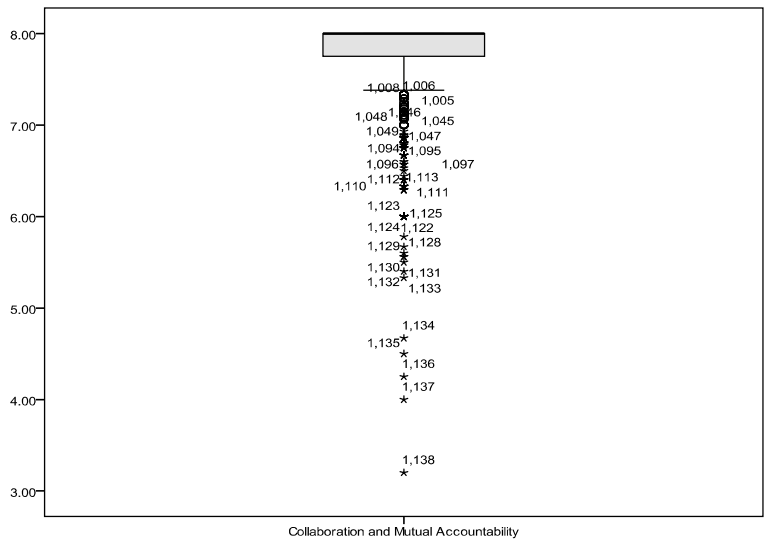


Figure 47: Box Plot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings.

To determine the type of statistical correlation to be conducted, a scatterplot was run to determine whether the variables were bivariately normally distributed or not. As can be seen in Figure 48, the scatterplot suggested that there was very little relationship between the variables, but there was not a non-linear relationship. Therefore, a Pearson Product-Moment Correlation was reasonable in being used to determine if a statistically significant correlation existed between the two variables. An alpha of .05 was used to control for the risk of making a Type I error, or the rejection of a true null hypothesis.

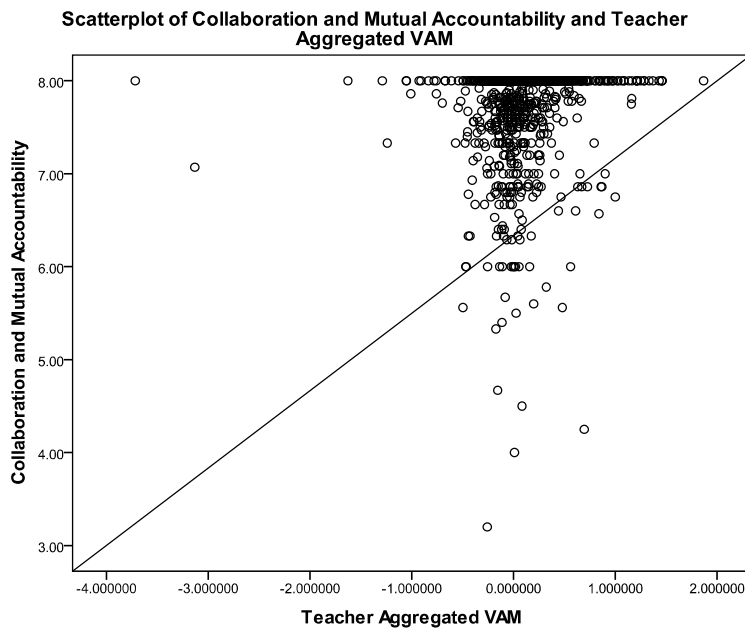


Figure 48: Scatterplot of Collaborative Inquiry (Collaboration and Mutual Accountability) Ratings.

The results of the correlational analysis indicated that there was a positive relationship between the collaborative inquiry (Collaboration and Mutual Accountability) and the VAM score,  $r(1,136) = .075$ ,  $r^2 = .01$ ,  $p = .011$ . Shared variance between these two variables was approximately 1%, generally interpreted to be a small effect (Cohen, 1988). Only 1% of the variability in VAM scores was accounted for by the collaborative inquiry (Collaboration and Mutual Accountability) ratings. When the correlation was

rerun with the two VAM extreme outliers and the five collaborative inquiry extreme outliers filtered out, there was still a weak, positive correlation between the two variables,  $r(1,129) = .089$ ,  $r^2 = .01$ ,  $p = .003$ . The data analysis confirmed a significant relationship between collaborative inquiry (Collaboration and Mutual Accountability) and VAM scores; and therefore, the null hypothesis was rejected.

### Summary

The population data used for all eight of the research questions in this study were not normally distributed. Although there was very little relationship with the VAM data as indicated by the scatterplots, the absence of a non-linear relationship supported the use of a Pearson Product-Moment Correlation as the statistical test in determining the correlation between the variables for this study. The correlation coefficients were consistent ranging from .089 to .231; and therefore, the presence of a small to moderate statistically significant, positive relationship between the variables and VAM scores existed. As a result, all null hypotheses were rejected. A summary of the study, discussion of the findings, implications for practice, recommendations for future research, and conclusions are presented in Chapter 5.

## CHAPTER V SUMMARY, DISCUSSION, AND CONCLUSIONS

This chapter contains a summary of the study, discussion of the findings, implications for practice, further recommendations for Brevard Public School District, recommendations for future research, and conclusions. The intent of this chapter is to provide a deeper understanding of the results of the research and their impact on determining which components of an evaluation system are more closely related to student learning as well as to provide recommendations for future research aimed at developing an evaluation instrument that reliably measures quality teaching.

### Summary of the Study

The purpose of this study was to determine the relationship between the professional practices components of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument and student learning as measured by a teacher's VAM score.

The evaluation instrument component ratings for Brevard Public Schools instructional personnel study participants were used for this research in addition to their VAM scores. The component ratings were taken directly from each teacher's Evaluation Summary Calculation Form that was completed by their evaluator and submitted to the Human Resource Department of the school district. The following constructs were measured by the components: professional practices; instructional practices; quality of instruction; professional responsibility, conduct and relationships; professional

development; action research development; action research implementation; and collaborative inquiry. The aggregated VAM scores were provided by the State Department of Education and were accessible for this study from the Testing and Accountability Department of the school district. A Pearson Product-Moment Correlation was the statistical method used to determine whether a relationship existed between the evaluation instrument components and student learning growth as measured by the VAM.

The study included 1,138 Brevard Public School fulltime instructional personnel (736 elementary, 177 middle school, 59 junior/senior high school, and 166 high school) in Grades 4-10 in which reading and/or mathematics was taught, the FCAT was administered, and a teacher aggregated VAM score was assigned. The population excluded the instructional personnel at charter schools, the adult education centers, alternative learning centers, virtual schools, and the school board department. An analysis of the demographics was presented for gender, ethnicity, and years of experience. This study was based on eight research questions:

1. To what extent, if any, is there a relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

$H_0$ : There is no relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

2. To what extent, if any, is there a relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

3. To what extent, if any, is there a relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

4. To what extent, if any, is there a relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

5. To what extent, if any, is there a relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

6. To what extent, if any, is there a relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

7. To what extent, if any, is there a relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

8. To what extent, if any, is there a relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?

H<sub>0</sub>: There is no relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10.

### Discussion of the Findings

This section examines the implications relative to the findings for each of the eight research questions. Some research questions have been grouped according to the components they reference so that discussions can be more meaningful.

Goodwin and Leech (2006) pointed out that certain characteristics of datasets such as range restriction and measurement error can affect the strength of a correlation. Data that have a restricted range underestimate the correlation. The ranges of the ratings for all of the variables, excluding the VAM scores, used in this study were small due to the nature of the instrumentation. Therefore, it can be anticipated that all of the correlations were underestimated based on range restriction. Additionally, measurement error is a factor that can also result in attenuation, or a lower correlation between two variables. The review of literature revealed that the reliability of VAM scores used in isolation has been determined to be low (Baker et al., 2010; Harris, 2012; Koedel &



Betts, 2007; Lockwood et al., 2002; McCaffrey et al., 2004). There is a correction for attenuation formula for estimating a correlation with measurement error eliminated from the variables; however, the reliability of the measures is required. Since the instruments used to gather data for this study were in their initial year of implementation, reliability measures had not been established. Based on range of restriction and attenuation, it is believed that the correlations in this study have been underestimated.

### *Research Question One*

*To what extent, if any, is there a relationship between professional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

The results of the data analysis for research question one indicate a significant but weak to moderate positive relationship between professional practices and VAM scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. As a teacher's overall evaluation rating increased, so did the value-added score and vice versa. This finding addresses the validity of the evaluation instrument in measuring teacher quality and the impact teacher quality has on student learning. It was comprised of the observation, professional growth plan, and collaboration and mutual accountability scores.

The cumulative evaluation ratings should parallel student academic growth (The New Teacher Project, 2010). Why the weak correlation? The professional practices scores were skewed to the right indicating that overall the scores were higher than a normal distribution. This means that the district is still experiencing to a lesser degree;

however, what Weisberg et al. (2009) called the Widget Effect where all teachers' classroom effectiveness was judged the same with no distinction being made between the good and the poor teachers. Within the population of this study ( $N = 1,138$ ), 854 (75%) subjects were considered "highly effective" according to their professional practices score; 273 (24%) were considered "effective" according to their overall score; and 11 (1%) were considered "needs improvement" according to their score. Therefore; 1,127 (99%) teachers were rated effective or highly effective; yet the value-added scores indicated that 494 (43%) of the teachers received negative value-added scores meaning that the collective progress of the students in these teachers' classrooms was less than expected. Because all of the components of the professional practices total rating were subjective measurements scored by trained administrators, colleagues, and teachers themselves, the degree to which they correlated with student learning was dependent upon the level of fidelity with which they were implemented.

A major implication of this finding is that no matter how reliable the components of an evaluation system are, it is not the instruments in and of themselves that determine effectiveness but rather how they are administered. Evaluator bias needs to be minimized if correlations are going to be strengthened. There is no neutrality in implementing an evaluation system when a relationship and previous knowledge exists between the evaluator and the teacher being evaluated. There is a humanistic driven inflation of scores or in some cases deflation of scores (McClellan, 2012). The real problem that needs to be addressed is how to implement an evaluation system in which bias can be minimized. According to McClellan, in order to minimize observer bias, training must distinguish between bias, interpretation, and evidence. Not only would these processes

need to be taught but they would also need to be assessed in order to gain certification. Although no new evaluation system is without flaws in its first year of implementation, it is necessary to gather data as to its effectiveness and make changes to the system and/or provide additional training to improve its reliability.

The correlation of the professional practices total score to VAM scores as seen in Table 1 was stronger than the correlation of any of its components. This supports the need for multiple data sources when documenting performance (Kane & Staiger, 2012; Kane et al. 2010; Peterson, 2004; Strong & Tucker, 2003).

Table 1: Professional Practices Correlation

Component	Correlation Coefficient
Professional Practices	.231
Instructional Practices	.209
Quality of Instruction	.218
Professional Responsibility, Conduct and Relationships	.188
Professional Development	.189
Action Research Development	.102
Action Research Implementation	.175
Collaborative Inquiry	.089

*Note. All correlations are significant at the .05 level.*

### *Research Question Two*

*To what extent, if any, is there a relationship between instructional practices and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

The results of the data analysis for research question two indicate a significant but weak to moderate positive relationship between instructional practices and VAM scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. As a teacher's observation rating increased, so did the VAM score and vice versa. This finding addresses the validity of the observation instrument in measuring teacher quality and the impact teacher quality has on student learning. It was comprised of the quality of instruction and professional responsibility, conduct and relationships scores.

As supported by the review of literature, classroom observation scores are related to the outcome measure of student learning with value-added correlations ranging from weak to moderate (Borman & Kimball, 2004; Gallagher, 2004; Jacob & Lefgren, 2008; Kane, Taylor, Tyler, & Wooten, 2010; Schacter & Thum, 2004). The research also indicated that to ensure accuracy, scores needed to be averaged across several lessons and conducted by multiple raters (Gallagher, 2004; Kane, Taylor, Tyler, & Wooten, 2010; Schacter & Thum, 2004; Stronge & Tucker, 2003). The January 2012 Measures of Effective Teaching (MET) report issued by the Bill and Melinda Gates Foundation found that "as the number of raters increased (inter-rater reliability) and the number of lessons observed increased (e.g., four different raters observing four different lessons), the higher

the reliability of the scores for any individual teacher” (Marzano, Toth, & Schooling, 2012, p. 19). A major implication is that in order to strengthen the correlation, multiple raters and multiple observations should be implemented. The evaluation system associated with this study required two formal observations for annual teachers and only one formal observation for tenured teachers. The use of multiple raters was not a requirement of this evaluation system.

Calibrating raters to the standards so that they can consistently rate teachers across numerous lessons improves the reliability of the instrument. This supports the Goals and Roles Evaluation Model theoretical framework upon which this study was based. Stronge and Tucker (2003) emphasized that the standards must be clear and easily understood by everyone involved in the evaluation process. The implication here is that observation ratings are based on judgments of good teaching, and training methods need to be developed for ensuring ratings that are consistent and reliable across all subject areas. The correlation between observation ratings and VAM scores is contingent on quality implementation which improves with intensive training (Kane & Staiger, 2012). This author agrees with Peterson (2004) who emphasized that the mere existence of behaviors as identified on a checklist or rubric do not equate with student learning; but rather the appropriateness of how these behaviors were performed. It is important that training clearly delineate between the various levels of performance and the appropriateness of their execution. For example, differentiating content is considered a practice of highly effective teachers; however, it is how a teacher differentiates the content according to student learning needs that makes the difference not the mere existence of differentiation. Analyzing evaluation data from the first-year of

implementation will enable the district to provide targeted support for evaluators based on their needs.

### *Research Question Three*

*To what extent, if any, is there a relationship between quality of instruction and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

The results of the data analysis for research question three indicate a significant but weak to moderate positive relationship between quality of instruction and VAM scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. As a teacher's quality of instruction observation rating increased, so did the VAM score and vice versa. This finding addresses the validity of the quality of instruction ratings in measuring teacher quality and the impact teacher quality has on student learning. It was comprised of Dimension 1 - Instructional Design and Lesson Planning, Dimension 2 – Learning Environment, Dimension 3 – Instructional Delivery and Facilitation, and Dimension 4 – Assessment.

A strong positive correlation existed between the quality of instruction observation ratings and the professional responsibility, conduct and relationships observation ratings (Research Question Four),  $r(1,136) = .80$ ,  $r^2 = .64$ ,  $p = .000$ . Teachers who were rated high on the quality of instruction dimensions of the classroom observation were also rated high on the professional responsibility, conduct and relationships dimensions of the classroom observation. In fact, 64% of the variance in

one rating was accounted for by the other rating and vice versa. These results could be interpreted as a pattern of raters scoring teachers consistently across all dimensions. This implies that further analysis is necessary to determine if positive or negative ratings in one dimension influence the rater's scores in other dimensions. If this is found to be true, then precautionary measures need to be implemented to eliminate this from happening.

Dr. Richard Elmore (City, Elmore, Fiarman, & Teitel, 2009) claims that the only way to increase learning is changing all three components of what he calls the instructional core: (a) teachers' knowledge and skill, (b) students' role in their own learning, and (c) level of content. The quality of instruction observation rating must focus on all three components. An implication of the results of this research question is that observers cannot restrict their ratings to only what the teacher is doing or the content expectations; they must also judge the quality of the performance task and determine the appropriateness of the task based on the level of student learning that results. The observation rubrics must be evaluated closely to assure that the elements are aligned with all three components of the instructional core and that they are not overemphasizing the important of one component as compared to the others. Also, observers must be trained to examine all three components when making judgments as to the level of proficiency for each element.

#### *Research Question Four*

*To what extent, if any, is there a relationship between professional responsibility, conduct and relationships and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

The results of the data analysis for research question four indicate a significant but weak to moderate positive relationship between professional responsibility, conduct and relationships and VAM scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. As a teacher's professional responsibility, conduct and relationships observation rating increased, so did the VAM score and vice versa. This finding addresses the validity of the professional responsibility, conduct and relationships ratings in measuring teacher quality and the impact teacher quality has on student learning. It was comprised of Dimension 5 – Professional Responsibilities and Ethical Conduct, Dimension 6 – Relationships with Students, and Dimension 7 – Relationships with Parents and Community Building Relationships with Parents.

If student and parent relationships are related to student learning, there must be more emphasis placed on the critical roles that parents and the community can play in impacting student learning. An implication is that it is crucial that the district and schools foster a partnership with parents and the community through communication, involvement activities, and the dissemination of knowledge as to what they can do to support learning. The role of the parent liaison for the district must be re-examined and clearly defined as how best to utilize the position to support parent involvement at the school level. Well planned partnerships between the parents, community, and schools linked to the goals of school improvement must be established. Additionally, school districts must investigate school reform models that lend themselves to a more positive learning environment and stronger relationships among teachers and students. According



to Montalvo (2007), there is a positive correlation between student-teacher relationships and student achievement.

*Research Question Five, Six, and Seven*

*Research Question 5: To what extent, if any, is there a relationship between professional development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

*Research Question 6: To what extent, if any, is there a relationship between action research development and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

*Research Question 7: To what extent, if any, is there a relationship between the implementation of action research and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

Professional development, action research development, and the implementation of action research are all components of this evaluation system that measure elements of the same construct; therefore, the findings relative to their analysis will be discussed together. The results of the data analysis for research questions five, six, and seven indicate a significant but weak to moderate positive relationship between professional development, including the development and implementation of the professional growth plan, and VAM scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. As a teacher's

professional growth plan development and implementation ratings increased, so did the VAM scores and vice versa. This finding addresses the validity of the professional growth plan, as well as the development and implementation of the plan in measuring teacher quality and the impact teacher quality has on student learning.

The development of the professional growth plan had one of the weaker correlations to VAM scores. The score a teacher receives for the professional growth plan development may be influenced by the teacher's ability to write. Although there was a significant positive correlation between the development and the implementation of the professional growth plan, it was considered weak,  $r(1,136) = .12$ ,  $r^2 = .01$ ,  $p = .000$ . Only 1% of the variance in the professional growth plan development rating was accounted for by the professional growth plan implementation rating and vice versa. One could conclude that there could be a well written professional development plan that is not fully implemented, and there could be a poorly written professional development plan that is fully implemented. As the literature review indicated, there is limited research on professional development plans and their correlation to VAM scores (exceptions include NGA Center for Best Practices, 2009; Wallace, 2009). There are numerous limitations to determining if a correlation exists between professional growth plans and value-added assessment scores. The annual plans are only implemented for a four-month period, so it is difficult to have a large impact on student learning in such a short period of time. Also, the learning targeted in the professional growth plan might not be what is being measured by the VAM score. For example, with the freedom to select an area of focus on which to base one's plan, if a teacher chose to concentrate on increasing writing scores, this would not correlate to the VAM scores that measure reading and mathematics learning only.

The implication here is that further research that controls for the limitations of the existing research relative to professional growth plans is necessary to determine the true correlation between professional growth plans and VAM scores.

#### *Research Question Eight*

*To what extent, if any, is there a relationship between collaborative inquiry and value-added assessment scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument at Grades 4-10?*

The results of the data analysis for research question eight indicate a significant positive but weak relationship between collaborative inquiry and VAM scores of instructional personnel on the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument. As a teacher's collaborative and mutual accountability rating increased, so did the VAM score and vice versa. This finding addresses the validity of the collaborative and mutual accountability ratings in measuring teacher quality and the impact teacher quality has on student learning.

The results of this study confirm the research on collaboration aimed at improving instructional practices. Research verified collaboration's positive effect on student learning (Goddard et al., 2007; Hollins et al., 2004; Phillips, 2003; Schmoker, 2004; Strahan, 2003; Supovitz, 2002). The weak correlational results could be a result of the narrow focus of the collaborative group effort as compared to the wide range of VAM scores. Whereas the collaborative group ratings were based on an individual's efforts at collaboratively targeting the lowest 25% students, this was correlated against the value-added scores of all students and not just the lowest 25%. The VAM scores for these

students would have to be isolated in order to determine if a stronger correlation existed. The implication here is that schools must be transformed into learning organizations by putting both formal and informal structures in place that promote a collaborative culture.

### Implications for Practice

Research suggests that the quality of teaching influences student learning (Haycock & Hanushek, 2010). In order to promote quality teaching, it is essential to have a valid evaluation instrument that identifies teacher competency and provides an avenue for support and improvement. This study's findings have major implications not only for the Brevard Public Schools, whose evaluation instrument was the focus of this study, but also for school districts and policy makers who are interested in improving learning for all students by means of evaluation instruments.

The purpose of the evaluation instrument under study was to determine its validity in measuring quality instructional performance. There was a significant but weak to moderate positive correlation between all of the components that comprised the professional practices portion of the instrument and VAM scores that represented student learning growth. This study confirmed that the Goals and Roles Evaluation Model (Stronge & Tucker, 2003) theoretical framework, that is in alignment with the framework upon which Brevard Public Schools' evaluation instrument was based, is valid in measuring effective teaching. It also magnifies the importance of ensuring that teachers receive high-quality feedback from their evaluations and the professional development and cognitive coaching needed to develop their deficiencies. Evaluation results should also be shared with instructional coaches who can further provide targeted support. A system should exist in school districts that align professional development opportunities

with all of the evaluation components on which the teacher is rated so that teachers can match workshops and course offerings to their targeted areas for growth.

Because correlations are based on the reliability of the measures, school districts must review the research evidence relative to each measure and structure each measure according to the most reliable outcomes. Since observational instruments are a major part of evaluation systems, it is important that school districts review the observation components to assure that they are comprehensive enough to provide a complete picture of effective teaching; however, not so extensive that they become unmanageable and decrease their reliability. They must also refine them by clarifying language, strengthening descriptors, and eliminating redundancy to increase the reliability of the instrument.

Additional measures as seen in this study increase the validity of an evaluation system, but at what point does the cost outweigh the additional gain? This is a decision that school districts and policy makers will need to consider in designing the requirements for an evaluation system using multiple measures of teacher performance.

For school district level personnel, this study reinforces the importance of implementation fidelity when dealing with evaluative measures based on human scoring. The Goals and Roles Model (Stronge & Tucker, 2003) used as the theoretical framework for this study emphasized standardization across the entire system. According to Danielson, “Training is the lynch-pin for assuring the new systems are successful” (Vevea, 2012, par 4). Evaluations that are accurate require evaluators with extensive knowledge and dedication. They need to know what constitutes good teaching and how to recognize it. Given the complexity of teaching and the range of grade levels and

subject areas, this is a huge undertaking. High quality, ongoing training is necessary for evaluations to be valid and impact student learning.

Evaluation systems must be structured in such a way to minimize observer bias. This might include multiple observations using observers with no previous knowledge of the teachers they are observing. Also, training should not only teach but assess an observer's ability to distinguish between bias, interpretation, and evidence. Additionally, videos should be made available for all observation components and demonstrate both the high end and low end of the performance range.

After the initial year of implementation of a new evaluation system, school level correlations must be conducted to identify those administrators needing additional training and their targeted area(s) of support. By using data results to norm the ratings, school districts can establish minimum standards that must be met by all evaluators before they are certified and permitted to carry out formal evaluations. This must include an assessment required of all evaluators' as to their ability to reliably rate teachers. These training modules must be judged according to their effectiveness and resources set aside to make improvement changes as necessary. A recertification process must also be built into the system on a periodic basis so that observers can maintain and improve their skills. According to Jerald (2012) training should be conducted in real classrooms because the transfer of skills to the building level is difficult for observers when training is based exclusively on recorded lessons.

Additionally, the entire evaluation system needs to be continuously scrutinized in light of new research and yearly data and modifications made accordingly. The need for inter-rater reliability suggests another implication for district level personnel in line with

what Stronge and Tucker (2003) recommend in the underlying theoretical framework for this study. An evaluation system must be based on multiple raters and multiple observations. Accordingly, structures must exist at the district level that enable personnel to monitor evaluation data from multiple evaluators so that inconsistencies can be identified and training aimed at better understanding the rubrics provided. Policy makers must recognize that a comprehensive evaluation system is more labor-intensive and requires funding to support its implementation. This is in direct conflict with the reduction of personnel presently taking place in many school districts.

As seen in this study, the correlation between teacher performance and student learning is weak. Policy makers need to determine at what level measures will be considered valid and reliable in order to minimize classification errors. With high stake decisions such as pay and retention tied to evaluation outcomes, it is imperative that teachers are not misclassified.

#### Further Recommendations for Brevard Public School District

Because this study confirms that there is a significant but weak to moderate positive correlation between the evaluation professional practices components of School Board of County Instructional Personnel Performance Appraisal System Instrument and student growth as measured by VAM scores, it is imperative that administrators and teachers be trained on how to use the evaluation data to identify areas for improvement in order to inform professional development. Both administrators and teachers need to have access to individual teacher and student VAM scores so that they can analyze the data for trends relative to their strengths and weaknesses. Administrators also need to be trained on how to triangulate the results of the evaluation components so that any problems

associated with the evaluation process can be identified and rectified. Ongoing professional development for both teachers and administrators must continue to be the focus of the district so that inter-rater reliability within schools and across schools continues to improve thus improving the validity of the instrument.

In order to increase the validity of the instrument, standards must be clearly defined, and both administrators and teachers must agree not only on their meaning but also on how they look within a classroom. Ongoing professional development aimed at calibrating the instrument is essential in defining what the standards look like and how they are differentiated by the various performance levels. Lesson videos exemplifying all of the observation elements should be available for all observation elements as well as the high and low performance levels within each element. Teachers that have been identified by their VAM scores as having had a negative impact on student growth should receive intensive training relative to the standards and how they are effectively implemented in the classroom. Training may include showcasing teachers who received high positive VAM scores. Likewise, administrators whose evaluation ratings do not align with their VAM results should receive the same intensive training as well as multiple opportunities to rate teachers and compare results to ratings received from multiple other observers. This type of required training could be developed online and written as a response-based program in which the sequence of the instruction is based on an individual's responses.

As this study confirmed, multiple measures are more reliable than any single component when evaluating teachers. Knowing this, it would seem logical that the same would hold true when using data sources to determine student growth. In addition to VAM scores, the district should use additional measures in determining student learning.



For example, pre- and posttests can be used as a source for measuring student learning as well as other forms of classroom evidence that reflects teaching effectiveness.

Finally, it is important that the school district seek feedback from both the teachers and the administrators relative to the various evaluation components. Further clarification and the elimination of duplicated construct measures should be discussed.

### Recommendations for Future Research

The purpose of this study was to determine the relationship between the professional practices components of the School Board of Brevard County Instructional Personnel Performance Appraisal System Instrument and student achievement as measured by a teacher's VAM score. The collection and analysis of data resulted in significant findings. However, there were limitations relative to this study that can be used as the basis for recommended future research.

Because this study focused directly on the evaluation system as implemented in the Brevard Public Schools, the same study could be replicated in other school districts to determine if the findings still hold true relative to their implementation. Additionally, because evaluation systems vary from school district to school district and state to state, this study can be replicated across other school districts whose evaluation components differ to determine which combination of measures most accurately reflects teaching effectiveness that results in student learning.

Future research could also include a multilevel model which would allow for an examination of how much variation there is between schools as well as within schools. This would enable the school district to uncover the overall patterns of difficulty as well as the individual weaknesses necessary to provide continued training. This analysis

would also make it possible for the district to identify schools with strong correlations and investigate what methods of implementation they utilized so that these could be shared amongst other schools. In addition, future research could include a further analysis of the data used in this study by correlating VAM scores with professional practices based on teacher, school, and observer demographics; grade level; and subject area. Any patterns revealed during the analysis would further enhance the school district's ability to target its training based on specific needs. Identifying teachers by quartiles based on their VAM scores and correlating these to their professional practice ratings would also provide input as to where observers were the most and least effective.

While not part of this study, a correlation was run to determine the relationship between professional practices and VAM scores based on building level. The alpha, or level of significance, was set at .05. At the elementary level, there was a positive but weak to moderate significant correlation,  $r(734) = .18, r^2 = .03, p = .000$ . At the middle school level there was a positive but weak to moderate significant correlation,  $r(175) = .26, r^2 = .06, p = .001$ . At the junior/senior high school level there was a positive but weak to moderate significant correlation,  $r(57) = .27, r^2 = .07, p = .043$ . At the high school level there was not a significant correlation,  $r(164) = .22, r^2 = .05, p = .096$ . Further analysis is necessary to determine why correlations were not statistically significant at the high school level. According to Tyler (2011), there are so many content areas at the high school level that being evaluated by someone with subject area expertise is impossible; and therefore, the evaluation results might lack reliability and validity.

In order to determine the effect that observer bias has on the evaluation instrument, future research might include correlating neutral party observation scores

with building-level observation scores. This study would provide input necessary to train and certify principals. It would also show how well observers have been trained in distinguishing between bias, interpretation, and evidence.

In order to judge the impact of the new evaluation system on student learning, future research would include a study that compares a teacher's VAM score from the previous evaluation system with their VAM score from the new evaluation system. Influencing factors such as grade level and subject would need to be held constant. Additionally, the reliability of the new measures could be determined by correlating the evaluation results from the first year of implementation to those from the second year for teachers in the same grade level and subject area.

As indicated in this study, one of the weaker correlations was collaborative inquiry with VAM scores. Because a teacher's collaboration and mutual accountability rating was based on the progress of the lowest 25% students, further research is necessary to identify these students' VAM scores and rerun the statistical test.

Another recommended study that became apparent when analyzing the data for this study was the misalignment between the PGP areas of focus and the subject areas being measured by the VAM scores. In order to determine whether or not a correlation existed between professional development and student learning, the PGPs that focused on reading and/or mathematics should be identified and correlated with their corresponding VAM scores. All other PGPs associated with other subject areas should be excluded.

An important consideration for any new evaluation system would be to determine whether or not its implementation was a sound decision with the benefits outweighing the costs. In other words, was the time and energy required to put it into practice worthy of

its outcomes? In order to get the answer to this question, further research might include a cost/benefit analysis to determine the feasibility of implementing the evaluation system with fidelity using the data presented in this study.

### Conclusions

This study broadened the knowledge base available from other researchers relative to teacher evaluation systems and what components are related to student learning as measured by VAM scores. According to the results of this study, a teacher's professional practices; instructional practices; quality of instruction; professional responsibility, conduct, and relationships; professional development; action research development; action research implementation; and collaborative inquiry are all evaluation system components that have a significant but weak to moderate positive correlation with a teacher's aggregated VAM scores. This study further confirmed the findings of current research and the Goals and Roles theoretical framework that multiple data sources are necessary when documenting a teacher's performance. It also confirmed the need for intensive training and a certification process with robust standards. Additionally, training must include a thorough understanding and assessment of bias, interpretation, and evidence. Only by providing observers with this level of training will the reliability of the instrument improve. With significant consequences being tied to evaluation results, it is imperative that the evaluation instrument results be accurate.

APPENDIX A  
PERMISSION TO REPRODUCE – DR. JAMES STRONGE

1183 Old Millpond Road  
Melbourne, FL 32940

March 12, 2012

Dr. James H. Stronge  
The College of William and Mary  
School of Education, P.O. Box 8795  
Williamsburg, VA 23187-8795

Dear Dr. Stronge,

I am completing a doctoral dissertation at the University of Central Florida entitled "An Analysis of the School Board of Brevard County Instructional Personnel Performance Appraisal Instrument and the Validity of Its Components in Measuring Student Achievement." I would like your permission to reprint in my dissertation your Goals and Roles Evaluation Model as presented on page 27 of your book entitled, Handbook on Teacher Evaluation: Assessing and Improving Performance. This model is being used as the theoretical framework for my dissertation.

The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own or your company owns the copyright to the above-described material.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you for your attention in this matter.

Sincerely,

*Carol A. Mela*

Carol A. Mela

---

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE:

By:   
Dr. James H. Stronge

Date: 3/21/12

*Best wishes with your study!  
JH*

APPENDIX B  
OVERVIEW OF SEARCH PARAMATERS

Components	Professional Growth Plans	Collaboration and Mutual Accountability	Value Added Assessment
<b>Defined Terms</b>	<ul style="list-style-type: none"> <li>• Action research</li> <li>• Classroom inquiry</li> <li>• Individual/Individualized growth plan</li> <li>• Job-embedded professional development</li> <li>• Practitioner inquiry</li> <li>• Practitioner research</li> <li>• Professional development</li> <li>• Professional growth plan</li> <li>• Results-based professional development</li> <li>• School improvement</li> <li>• Self-study</li> <li>• Staff development</li> <li>• Teacher research</li> <li>• Teaching improvement</li> </ul>	<ul style="list-style-type: none"> <li>• Collaboration</li> <li>• Collective commitment</li> <li>• Educational cooperation</li> <li>• Mutual accountability</li> <li>• Professional community</li> <li>• Professional learning community</li> <li>• School culture</li> <li>• Shared vision</li> <li>• Teacher collaboration</li> <li>• Teacher interaction</li> <li>• Teacher teams</li> <li>• Teachers working together</li> </ul>	<ul style="list-style-type: none"> <li>• Accountability</li> <li>• Measuring teacher effectiveness</li> <li>• Race to the Top</li> <li>• Senate Bill 736</li> <li>• Student achievement</li> <li>• Student learning</li> <li>• Student outcomes</li> <li>• Teacher evaluation</li> <li>• Value-added assessment</li> <li>• Value-added models</li> <li>• Value-added scores</li> <li>• VAM</li> </ul>
<b>Primary Researchers</b>	<ul style="list-style-type: none"> <li>• Andrew Croft</li> <li>• Barbara Tye</li> <li>• Christine Wahlquist</li> <li>• David Dunaway</li> <li>• David Marsh</li> <li>• Dennis Sparks</li> <li>• Elizabeth Powers</li> <li>• Elizabeth Szad</li> <li>• Jackie Thompson</li> <li>• Joan Barnatt</li> <li>• Joellen Killion</li> <li>• John Bransford</li> <li>• John Holloway</li> <li>• Kathie Bone</li> <li>• Kay Burke</li> <li>• Kaye Chatterton</li> <li>• Kenneth Peterson</li> <li>• Kenneth Tye</li> <li>• Kim Do-Hong</li> <li>• Linda Darling-Hammond</li> <li>• Marilyn Cochran-Smith</li> <li>• Marjorie Wallace</li> <li>• Megan Dolan</li> <li>• Michael Fullan</li> <li>• Milbrey McLaughlan</li> <li>• Susan Loucks-Horsley</li> <li>• Susan Lyte</li> <li>• Thomas Guskey</li> </ul>	<ul style="list-style-type: none"> <li>• Agnes McMahon</li> <li>• Arthurlene Towner</li> <li>• Carl Rosenholtz</li> <li>• Charles DeBose</li> <li>• David Strahan</li> <li>• Douglas Smith</li> <li>• Edward Britton</li> <li>• Elizabeth Yeager</li> <li>• Etta Hollins</li> <li>• Helen Marks</li> <li>• Jon Katzenbach</li> <li>• Jonathan Supovitz</li> <li>• Joy Phillips</li> <li>• Karen Louis</li> <li>• Kathleen Fulton</li> <li>• Kimberly Honnins</li> <li>• Linda McIntyre</li> <li>• Louise Stoll</li> <li>• Lucretia Pence</li> <li>• Mary Brownell</li> <li>• Mary Rennells</li> <li>• Megan Tschannen-Moran</li> <li>• Mike Schmoker</li> <li>• Mike Wallace</li> <li>• Ray Bolam</li> <li>• Rick DuFour</li> <li>• Roger Godard</li> <li>• Sally Thomas</li> <li>• Sandra Musanti</li> <li>• Sharon Kruse</li> <li>• Steve Gruenert</li> <li>• Tamar Riley</li> <li>• Yvonne Goddard</li> </ul>	<ul style="list-style-type: none"> <li>• Andrew Beveridge</li> <li>• Cory Koedel</li> <li>• Daniel Koretz</li> <li>• Daniel McCaffrey</li> <li>• Diane Ravitch</li> <li>• Douglass Harris</li> <li>• Edward Haertel</li> <li>• Eva Baker</li> <li>• Ewart Thomas</li> <li>• Heather Hill</li> <li>• Helen Ladd</li> <li>• Henry Braum</li> <li>• J. R. Lockwood</li> <li>• Jennifer Jennings</li> <li>• John Papay</li> <li>• Julian Betts</li> <li>• Jun Ishii</li> <li>• Kata Mihaly</li> <li>• Kristin Umland</li> <li>• Laura Hamilton</li> <li>• Laura Kapitula</li> <li>• Linda Darling-Hammond</li> <li>• Lorrie Shepard</li> <li>• Paul Barton</li> <li>• Richard Rothstein</li> <li>• Richard Shovelson</li> <li>• Robert Gordon</li> <li>• Robert Linn</li> <li>• Sean Corcoran</li> <li>• Steven Rivkin</li> <li>• Thomas Louis</li> <li>• Tim Sass</li> <li>• William Sanders</li> <li>• Xiaoxia Newton</li> </ul>



Components	Teacher Quality	Historical Perspective		Observation	
<p><b>Defined Terms</b></p>	<ul style="list-style-type: none"> <li>• Effective teacher components</li> <li>• Effective teachers</li> <li>• Quality teachers</li> <li>• Teacher characteristics</li> <li>• Teacher classroom practices</li> <li>• Teacher effectiveness</li> <li>• Teacher performance</li> <li>• Teacher qualifications</li> <li>• Teacher quality</li> <li>• Teacher quality measures</li> </ul>	<ul style="list-style-type: none"> <li>• Changes in teacher evaluation</li> <li>• Evaluation methods</li> <li>• History of teacher evaluation</li> <li>• Teacher evaluation</li> <li>• Teacher evaluation history</li> <li>• Teacher evaluation in the past</li> <li>• Teacher evaluation models</li> <li>• Teacher evaluation timeline</li> <li>• Teacher evaluation trends</li> </ul>		<ul style="list-style-type: none"> <li>• Classroom observations</li> <li>• Effective classroom practices</li> <li>• Performance-based</li> <li>• Principal ratings</li> <li>• Student achievement</li> <li>• Student achievement growth</li> <li>• Teacher behavior</li> <li>• Teacher effectiveness</li> <li>• Teacher evaluation</li> <li>• Teacher observation</li> <li>• Teacher Observation validity</li> <li>• Teacher performance</li> <li>• Value-added assessment</li> </ul>	
<p><b>Primary Researchers</b></p>	<ul style="list-style-type: none"> <li>• Allen Odden</li> <li>• Anthony Milanowski</li> <li>• Courtney Bell</li> <li>• Florence Chang</li> <li>• Herbert Heneman</li> <li>• Laura Goe</li> <li>• Marco Munoz</li> <li>• Michael Strong</li> <li>• Olivia Little</li> <li>• Patricia Hinchey</li> <li>• Steven Kimball</li> </ul>	<ul style="list-style-type: none"> <li>• Andrea Touhey</li> <li>• C. W. Hill</li> <li>• Charlotte Danielson</li> <li>• Daisy Rucinski</li> <li>• Daniel Weisberg</li> <li>• David Keeling</li> <li>• Donald Medley</li> <li>• Erin Borthwick</li> <li>• Glenn Daley</li> <li>• Helen Hazi</li> <li>• Homer Coker</li> <li>• James Sennette</li> <li>• Jennifer Mulhern</li> <li>• Liam Goldrick</li> <li>• Lydia Kirr</li> <li>• Morgaen Donaldson</li> <li>• Robert Soar</li> <li>• Sarah Cohodes</li> </ul>	<ul style="list-style-type: none"> <li>• Susan Sexton</li> <li>• Thomas McGreal</li> <li>• William Connor</li> </ul>	<ul style="list-style-type: none"> <li>• Amy Wooten</li> <li>• Anthony Milanowski</li> <li>• Brian Jacob</li> <li>• Charlotte Danielson</li> <li>• Douglas Staiger</li> <li>• Douglas Staiger</li> <li>• Eric Hanushek</li> <li>• Eric Taylor</li> <li>• Geoffrey Borman</li> <li>• Hersh Waxman</li> <li>• James Stronge</li> <li>• John Schacter</li> <li>• John Tyler</li> <li>• Jonah Rockoff</li> <li>• Kenneth Aunchman</li> <li>• Lars Lefgren</li> <li>• Laura Goe</li> <li>• Linda Darling-Hammond</li> </ul>	<ul style="list-style-type: none"> <li>• Morgaen Donaldson</li> <li>• Sara Pease</li> <li>• Sharon Gallagher</li> <li>• Steven Kimball</li> <li>• Thomas Kane</li> <li>• Thomas McGreal</li> <li>• Yeow Meng Thum</li> </ul>

APPENDIX C  
PERMISSION TO REPRODUCE – DR. LAURA GOE

1183 Old Millpond Road  
Melbourne, FL 32940

March 11, 2012

Dear Dr. Goe,

I am completing a doctoral dissertation at the University of Central Florida entitled "An Analysis of the School Board of Brevard County Instructional Personnel Performance Appraisal Instrument and the Validity of Its Components in Measuring Student Achievement." I would like your permission to reprint in my dissertation your teacher quality framework flowchart that defines teacher quality as a combination of inputs, processes, and outcomes. Your research is being used as part of my literature review.

The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own or your company owns the copyright to the above-described material.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me. Thank you for your attention in this matter.

Sincerely,



Carol A. Mela

---

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE:



Laura Goe, Ph.D.

Date: March 11, 2012

APPENDIX D  
PERMISSION TO REPRODUCE – DR. DEBRA PACE

ASSOCIATE SUPERINTENDENT  
MAR 20 2012

1183 Old Millpond Road  
Melbourne, FL 32940

March 12, 2012

Ms. Debra Pace, Associate Superintendent  
Office of Human Resources Services  
Brevard County Public Schools  
2700 Judge Fran Jamieson Way  
Viera, FL 32940

Dear Ms. Pace,

I am completing a doctoral dissertation at the University of Central Florida entitled "An Analysis of the School Board of Brevard County Instructional Personnel Performance Appraisal Instrument and the Validity of Its Components in Measuring Student Achievement." I would like your permission to reprint in my dissertation the following documents and excerpts taken from Brevard Public School's IPPAS Handbook Human Resource website:

- The Performance Appraisal Model (p. 17)
- The Continuous Improvement Cycle (Appendix 1)
- Evaluation Summary Calculation Form (Appendix 42-43)
- Classroom Observation Instrument (Appendix 24-27)
- Observation Rubrics (Appendix 5-18)
- BPS Instructional Personnel Evaluation Instrument (Appendix 35-38)
- Teacher Professional Growth Plan (PGP) Goal
- Individual Professional Growth Plan (PGP) Development (Appendix 19-20)
- Professional Growth Plan (PGP) Development Scoring and Feedback (Appendix 32)
- Professional Growth Plan (PGP) Development Administrator Final Scoring and Feedback
- Mid-Year Conference
- Individual Professional Growth Plan (PGP) Implementation (Appendix 21)
- Professional Growth Plan (PGP) Implementation Scoring and Feedback (Appendix 34)
- Collaboration & Mutual Accountability Team Scoring Form
- Value-Added Measures (Appendix 52)

The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own or your company owns the copyright to the above-described material.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you for your attention in this matter.

Sincerely,

*Carol A. Mela*

Carol A. Mela

---

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE:

By: *DR*  
*Ms. Debra Pace*

Date: *4/4/12*

APPENDIX E  
COURSE LIST

FCAT Tested Courses Grades 4-10

Course #	Course Title	Statewide Standardized Assessment
0500960	Reading and Writing Across The Curriculum	FCAT 2.0
1000000	M/J Intensive Language Arts (MC)	FCAT 2.0
1000010	M/J Intensive Reading (MC)	FCAT 2.0
1000400	Intensive Language Arts	FCAT 2.0
1000410	Intensive Reading	FCAT 2.0
1001010	M/J Language Arts 1	FCAT 2.0
1001020	M/J Language Arts 1, Advanced	FCAT 2.0
1001030	M/J Language Arts 1 International Baccalaureate	FCAT 2.0
1001040	M/J Language Arts 2	FCAT 2.0
1001050	M/J Language Arts 2, Advanced	FCAT 2.0
1001060	M/J Language Arts 2, International Baccalaureate	FCAT 2.0
1001070	M/J Language Arts 3	FCAT 2.0
1001080	M/J Language Arts 3, Advanced	FCAT 2.0
1001090	M/J Language Arts 3, International Baccalaureate	FCAT 2.0
1001300	English Skills 1	FCAT 2.0
1001310	English 1 (Class including LEP student(s) must use certs	FCAT 2.0
1001320	English Honors 1	FCAT 2.0
1001330	English Skills 2	FCAT 2.0
1001340	English 2 (Class including LEP student(s) must use certs	FCAT 2.0
1001350	English Honors 2	FCAT 2.0
1001360	English Skills 3	FCAT 2.0
1001370	English 3 (Class including LEP student(s) must use certs	FCAT 2.0
1001380	English Honors 3	FCAT 2.0
1001390	English Skills 4	FCAT 2.0
1001400	English 4 (Class including LEP student(s) must use certs	FCAT 2.0
1001410	English Honors 4	FCAT 2.0
1001440	Business English 1	FCAT 2.0
1001450	Business English 2	FCAT 2.0
1001560	Pre-AICE English Language	FCAT 2.0
1001800	English 1-Preinternational Baccalaureate	FCAT 2.0
1001810	English 2-Preinternational Baccalaureate	FCAT 2.0
1002000	M/J Language Arts 1 Through ESOL	FCAT 2.0
1002010	M/J Language Arts 2 Through ESOL	FCAT 2.0
1002020	M/J Language Arts 3 Through ESOL	FCAT 2.0
1002180	M/J Developmental Language Arts Through ESOL (MC)	FCAT 2.0
1002300	English 1 Through ESOL	FCAT 2.0
1002310	English 2 Through ESOL	FCAT 2.0
1002320	English 3 Through ESOL	FCAT 2.0
1002380	Developmental Language Arts Through ESOL (MC)	FCAT 2.0
1002520	English 4 Through ESOL	FCAT 2.0
1005375	AICE English Literature 2	FCAT 2.0
1008010	M/J Reading 1	FCAT 2.0



<b>Course #</b>	<b>Course Title</b>	<b>Statewide Standardized Assessment</b>
1008020	M/J Reading 1, Advanced	FCAT 2.0
1008040	M/J Reading 2	FCAT 2.0
1008050	M/J Reading 2, Advanced	FCAT 2.0
1008070	M/J Reading 3	FCAT 2.0
1008080	M/J Reading 3, Advanced	FCAT 2.0
1008300	Reading 1	FCAT 2.0
1008310	Reading 2	FCAT 2.0
1008320	Advanced Reading	FCAT 2.0
1008330	Reading 3	FCAT 2.0
1205010	M/J Mathematics 1	FCAT 2.0
1205020	M/J Mathematics 1, Advanced	FCAT 2.0
1205040	M/J Mathematics 2	FCAT 2.0
1205050	M/J Mathematics 2, Advanced	FCAT 2.0
1205070	M/J Mathematics 3	FCAT 2.0
1205080	M/J Mathematics 3, Advanced	FCAT 2.0
2002100	M/J Comprehensive Science 3	FCAT 2.0
2002110	M/J Comprehensive Science 3, Advanced	FCAT 2.0
2400000	Sixth Grade	FCAT 2.0
5010010	English for Speakers of Other Language-Elementary	FCAT 2.0
5010020	Functional Basic Skills in Reading-Elementary	FCAT 2.0
5010045	Language Arts-Grade Four	FCAT 2.0
5010046	Language Arts-Grade Five	FCAT 2.0
5010050	Reading-Elementary	FCAT 2.0
5010060	Integrated Language Arts - Elementary	FCAT 2.0
5012060	Mathematics - Grade Four	FCAT 2.0
5012070	Mathematics - Grade Five	FCAT 2.0
5020060	Science - Grade Five	FCAT 2.0
7755010	Academics: K-5	FCAT 2.0
7810010	Language Arts: 6-8	FCAT 2.0
7810020	Reading: 6-8	FCAT 2.0
7910100	Reading: 9-12	FCAT 2.0
7910110	English: 9-12	FCAT 2.0
7910400	Life Skills Reading: 9-12	FCAT 2.0
1008010	M/J Reading 1	FCAT 2.0

APPENDIX F  
EVALUATION SUMMARY CALCULATION FORM



Evaluation Summary Calculation Form

Name \_\_\_\_\_ School Year \_\_\_\_\_
Last First MI
School/Dept Name \_\_\_\_\_ School/Dept # \_\_\_\_\_ Contract Status: INTERIM [ ] ANNUAL [ ]
Principal/Dept Head \_\_\_\_\_ Assignment \_\_\_\_\_ New Hire [ ] AC [ ] CC/PSC [ ]

Formative Evaluation Observation Component table with columns: Description, Pts Possible, Self Score, Pts Earned. Rows include Quality of Instruction (Dimensions 1-4) and Professional Responsibility, Conduct and Relationships (Dimensions 5-7).

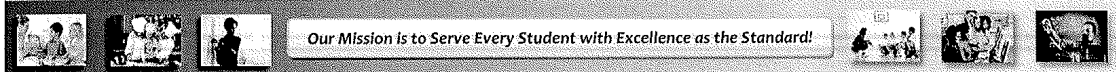
Continuous Professional Improvement Component table with columns: Description, Teacher Leader Scores (1, 2, 3, Average), Admin Score, Averaged Pts Earned. Rows include Professional Growth Plan (PGP) Development, PGP Implementation, and Collaboration and Mutual Accountability.

My signature indicates that this evaluation has been discussed with me:

Signature of Employee (Blue Ink Only) \_\_\_\_\_ / \_\_\_\_\_ Date

Signature of Principal/Administrative Supervisor (Required) (Blue Ink Only) \_\_\_\_\_ / \_\_\_\_\_ Date

IPPAS Appendix: 42





Evaluation Summary Calculation Form (continued)

Name \_\_\_\_\_  
Last First MI

School Year \_\_\_\_\_

Student Performance/Growth Measures		
Component 3: Description	Pts Possible	Pts Earned
1. Student Achievement Results		
1a. Value Added	35	
1b. Collaborative Team	5	
2. District Accountability Measures		
2a. School Improvement Plan	5	
i. Implementation		
ii. Outcome Indicators		
2b. Professional Growth Plan Outcomes	2	
2c. Whole school measured by regression of FCAT components of School Grade	3	
<b>Component 3: Cumulative Total (Sections 1 + 2)</b>	<b>50</b>	

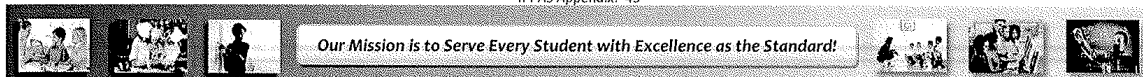
Alignment Component and Final Calculation		
Final Calculation	Pts Possible	Pts Earned
Formative Evaluation Total (Components 1 & 2)	47	
Student Performance Total	50	
Alignment Total (see rubric)	3	
<b>SUMMATIVE INSTRUCTIONAL PERFORMANCE APPRAISAL SYSTEM FINAL SCORE</b>	<b>100</b>	

Highly Effective (86-100 pts)     Effective (73-85 pts)     Needs to Improve (64-72 pts)     Unsatisfactory (63 - 0 pts)

My signature indicates that this evaluation has been discussed with me:

\_\_\_\_\_/\_\_\_\_\_  
Signature of Employee (Blue Ink Only) Date

\_\_\_\_\_/\_\_\_\_\_  
Signature of Principal/Administrative Supervisor (Required) (Blue Ink Only) Date



APPENDIX G  
IRB REVIEW AND BREVARD PUBLIC SCHOOLS RESEARCH  
APPROVAL



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

### Approval of Exempt Human Research

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

To: **Carol A. Mela**

Date: **March 19, 2012**

Dear Researcher:

On 3/19/2012, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review: Exempt Determination  
Project Title: An Analysis of the School Board of Brevard County Instructional Personnel Performance Appraisal Instrument and the Validity of its components in Measuring Student Growth  
Investigator: Carol A. Mela  
IRB Number: SBE-11-07931  
Funding Agency:  
Grant Title:  
Research ID: N/A

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

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

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

Signature applied by Joanne Muratori on 03/19/2012 09:24:18 AM EST

A handwritten signature in cursive script that reads "Joanne Muratori".

IRB Coordinator

April 17, 2012

Dear Mrs. Mela,

Thank you for your application to conduct research in the Brevard Public Schools. This letter is official verification that your application has been accepted and approved through the Office of Accountability, Testing, & Evaluation. However, approval from this office does not obligate the principal of the schools you have selected to participate in the proposed research. Please contact the principals of the impacted schools in order to obtain their approval. Upon the completion of your research, submit your findings to our office. If we can be of further assistance, do not hesitate to contact our office.

Sincerely,

*Vickie B. Hickey*

Vickie B. Hickey, Resource Teacher  
Office of Accountability, Testing, and Evaluation

APPENDIX H  
CLASSROOM OBSERVATION INSTRUMENT



## Classroom Observation Instrument

TEACHER \_\_\_\_\_ OBSERVER \_\_\_\_\_

This instrument is to be used for data collection through observation of classroom teaching. Dimensions that can be observed in the classroom comprise the elements to be observed

<b>DIMENSION 2: LEARNING ENVIRONMENT</b>				
<i>ESSENTIAL QUESTION:</i> What will I do to provide a learning environment that meets students' needs?	OBS	COMMENTS	DATE	TIME
1. Creates and maintains a safe and organized learning environment.				
2. Promotes a flexible, inclusive, collaborative, and student-centered learning environment.				
3. Allocates and manages time, space, and resources.				
4. Manages student conduct.				
5. Conveys high expectations, respects students' cultural backgrounds and accommodates diverse needs of students.				
6. Models and teaches clear, acceptable communication skills.				
7. Maintains a climate of inquiry.				
8. Integrates appropriate technologies.				

## Classroom Observation Instrument

TEACHER \_\_\_\_\_ OBSERVER \_\_\_\_\_

This instrument is to be used for data collection through observation of classroom teaching. Dimensions that can be observed in the classroom comprise the elements to be observed

<b>DIMENSION 3: INSTRUCTIONAL DELIVERY AND FACILITATION</b>				
<i><b>ESSENTIAL QUESTION:</b></i> What will I do to provide effective delivery of instruction to students?	OBS	COMMENTS	DATE	TIME
1. Utilizes knowledge of the subject to enrich students' understanding, identifies gaps in students' learning, and modifies instruction to respond to student misconceptions.				
2. Employs higher order questions.				
3. Applies varied instructional strategies and resources.				
4. Delivers engaging, challenging and relevant lessons.				
5. Differentiates instruction.				
6. Provides immediate and specific feedback to students.				

## Classroom Observation Instrument

TEACHER \_\_\_\_\_ OBSERVER \_\_\_\_\_

This instrument is to be used for data collection through observation of classroom teaching. Dimensions that can be observed in the classroom comprise the elements to be observed

<b>DIMENSION 4: ASSESSMENT</b>				
<i><b>ESSENTIAL QUESTION:</b></i>	<b>OBS</b>	<b>COMMENTS</b>	<b>DATE</b>	<b>TIME</b>
<b>What will I do to assess student learning?</b>				
1. Analyzes and applies data from multiple measures to diagnose students' learning needs, inform instruction, and monitor progress.				
2. Designs and uses formative and summative assessments that lead to mastery.				
3. Modifies teacher made assessments to accommodate diversity.				
4. Communicates assessment data to students and parents.				

## Classroom Observation Instrument

TEACHER \_\_\_\_\_ OBSERVER \_\_\_\_\_

This instrument is to be used for data collection through observation of classroom teaching. Dimensions that can be observed in the classroom comprise the elements to be observed

<b>DIMENSION 6: RELATIONSHIP WITH STUDENTS</b>				
<i><b>ESSENTIAL QUESTION:</b></i> How will I build relationships with students?	OBS	COMMENTS	DATE	TIME
1. Demonstrates knowledge of students.				
2. Builds relationships through instructional interactions.				
3. Creates a positive environment of respect and rapport.				

APPENDIX I  
OBSERVATION RUBRICS

**Dimension 1: Instructional Design and Lesson Planning**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Sets instructional outcomes and aligns instruction with state-adopted standards</b>	<ul style="list-style-type: none"> <li>Instructional goals and objectives are consistently clear, focus on student learning and achievement of standards and the end result is complete and coherent; directions, procedures, and content are appropriate for and clear to all students.</li> <li>Plans routinely to provide for instruction to meet the needs of students with varied levels.</li> </ul>	<ul style="list-style-type: none"> <li>Lesson plans are aligned to the district adopted curriculum maps and district/state assessments.</li> </ul>	<ul style="list-style-type: none"> <li>Lesson plans are not consistently aligned to the district adopted curriculum maps and district/state assessments.</li> </ul>	<ul style="list-style-type: none"> <li>Lesson plans, pacing guides, when available, do not align with the district adopted curriculum maps and/or district/state assessments.</li> </ul>
<b>Designs instruction using student prior knowledge and diagnostic student data to build coherent lessons</b>	<ul style="list-style-type: none"> <li>Regularly accesses and systematically uses data at the item level to find strengths and challenges both for disaggregated groups and for individual students.</li> <li>Uses these data to accurately refine and modify instruction for whole group, small groups and for specific individuals.</li> </ul>	<ul style="list-style-type: none"> <li>Makes accurate use of student achievement data and uses prior student knowledge when making instructional decisions.</li> <li>Examines data at the item level to find strengths and challenges for disaggregated groups of students to appropriately plan for instruction for whole and small groups.</li> </ul>	<ul style="list-style-type: none"> <li>Accesses data to view class achievement and examines data but shows little application to make instructional decisions linked to analysis of data.</li> <li>Inferences about the data may not be complete. Examines data at the group level and uses these data when planning instruction.</li> </ul>	<ul style="list-style-type: none"> <li>Relies on someone else to access student achievement data.</li> <li>When data is accessed, it may not be used to make decisions.</li> <li>Does not consider data to make changes in instruction.</li> </ul>
<b>Designs ways to monitor learning, designs use of appropriate instructional resources, and designs formative assessments</b>	<ul style="list-style-type: none"> <li>High quality classroom formative and summative assessments are designed that accurately measure student learning of the lesson or unit objectives.</li> </ul>	<ul style="list-style-type: none"> <li>Lesson or unit is planned in detail around clearly defined lesson objectives.</li> <li>Progression and pacing of the planned learning time (<i>instructional strategies, student activities, use of resources, assessment tasks</i>) are constant, with reasonable time allocations.</li> </ul>	<ul style="list-style-type: none"> <li>Lesson plans or units are based on activities or resources rather than focused on objectives.</li> <li>Progression and pacing of learning activities are sporadic, thus, time allocations are not always reasonable.</li> <li>Lesson plans show little evidence of formative assessment.</li> </ul>	<ul style="list-style-type: none"> <li>There is little evidence of monitored learning or use of appropriate instructional resources.</li> <li>No evidence of planning and design of formative assessment or progress monitoring.</li> </ul>
<b>Requires students to understand and demonstrate skills and competencies</b>	<ul style="list-style-type: none"> <li>Assessment tasks are embedded within the lesson and/or require a performance component for students to demonstrate what they know and are able to do.</li> </ul>	<ul style="list-style-type: none"> <li>Demonstration of skills and competencies is developed in advance of instruction and is aligned with the lesson objective.</li> <li>Lesson objectives are written in student-friendly language and students understand what they are expected to know and be able to do by the end of each lesson.</li> </ul>	<ul style="list-style-type: none"> <li>Although students are aware of the posted lesson objective, they rely on teacher direction to focus them on what they are expected to know and be able to do.</li> <li>The demonstration of skills and competencies is minimally developed and/or may be loosely connected to the lesson objective.</li> </ul>	<ul style="list-style-type: none"> <li>Lesson objectives are not known to students and students do not know what they are expected to know and be able to do.</li> </ul>

**Dimension 2: Learning Environment**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Creates and maintains a safe and organized learning environment</b>	<ul style="list-style-type: none"> <li>Implements and routinely reinforces, and students assume responsibility for school behavioral rules and regulations and established procedures to ensure student safety needs are consistently met.</li> <li>Establishes and maintains a comfortable, safe, and inviting learning environment that is organized to facilitate a focus on learning.</li> </ul>	<ul style="list-style-type: none"> <li>Implements and routinely reinforces school behavioral rules and regulations and established procedures to ensure student safety needs are consistently met.</li> <li>Implements classroom routines.</li> </ul>	<ul style="list-style-type: none"> <li>Does not consistently reinforce school behavioral rules and regulations and/or is not clear about established procedures in emergency situations.</li> <li>Lacks an appropriate level of attention to the supervision of students inside and outside the classroom.</li> </ul>	<ul style="list-style-type: none"> <li>Allows for conflict in the classroom environment.</li> <li>Is non-compliant and/or negligent with regard to supervision of students inside and outside the classroom setting.</li> </ul>
<b>Promotes a flexible, inclusive, collaborative, and student-centered learning environment</b>	<ul style="list-style-type: none"> <li>Fosters and creates, through collaboration among students, a stimulating and inclusive learning environment; maximizes potential for equal learning opportunities for every student.</li> <li>Teachers create an environment where students assume responsibility and hold themselves accountable for their learning.</li> <li>Instructional time is protected by students' adherence to classroom procedures.</li> </ul>	<ul style="list-style-type: none"> <li>Fosters a stimulating and inclusive learning environment.</li> <li>Teachers assume responsibility for student learning.</li> <li>Classroom procedures have been established and are followed by students.</li> </ul>	<ul style="list-style-type: none"> <li>Allows for too many students to "fall through the cracks".</li> <li>Does not individualize student attainment of standards.</li> <li>In the absence of known procedures, students spend time waiting for teacher re-direction in order to engage in independent work or cooperative group work.</li> </ul>	<ul style="list-style-type: none"> <li>Fails to use procedures for independent and/or cooperative group-work.</li> <li>Engages in learning solely on textbook or curriculum sequence with little or no individualized instruction for students to meet specific learning needs.</li> </ul>
<b>Allocates and manages time, space, resources</b>	<ul style="list-style-type: none"> <li>Arranges physical resources proactively to assure accessibility to all students, taking into account individual needs of students.</li> <li>Learning experiences, activities and physical spaces are highly organized and efficiently facilitated by both the teacher and students who each assume responsibility for maximizing time for learning.</li> </ul>	<ul style="list-style-type: none"> <li>Use of physical resources and space (furniture, technology, and learning stations) contributes to all students' ability to access learning.</li> <li>Learning experiences, activities and physical spaces are organized and facilitated by the teacher in such a way that students have adequate time for learning.</li> </ul>	<ul style="list-style-type: none"> <li>Use of physical resources and space limits student engagement and does not promote learning activities' accessibility to all students resulting in loss of learning time for some students.</li> <li>Failure to utilize and maintain facilities/equipment results in loss of learning opportunities.</li> </ul>	<ul style="list-style-type: none"> <li>Poor use of physical resources and space is evident and/or learning is not accessible to some students resulting in considerable "down-time."</li> <li>Teaching/ learning time wasted.</li> </ul>

**Dimension 2: Learning Environment**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Manages student conduct</b>	<ul style="list-style-type: none"> <li>Students contribute to designing the classroom rules and standards of conduct upheld by all members of the classroom.</li> <li>Consequences for inappropriate behavior are reasonable, clear, and consistently applied.</li> <li>Monitoring by the teacher of events in the classroom is subtle and proactive.</li> <li>Students monitor their own and their peers' behavior, correcting one another respectfully.</li> <li>The focus of discipline is to maximize student learning time and students show a respect for the rights of other students to learn.</li> </ul>	<ul style="list-style-type: none"> <li>Standards for student conduct are evident, clear, and the physical environment supports a nurturing learning environment.</li> <li>Teacher is alert to student behavior at all times and manages student conduct.</li> <li>Consequences for inappropriate behavior are reasonable, clear, and consistently applied.</li> </ul>	<ul style="list-style-type: none"> <li>Most students seem to understand rules of conduct.</li> <li>Teacher is generally aware but misses the activities of some students.</li> <li>Rules and consequences for inappropriate behavior are inconsistently applied.</li> <li>Management techniques do not consistently assist students in developing individual responsibility for learning and behavior.</li> </ul>	<ul style="list-style-type: none"> <li>Classroom rules and consequences have not been established, communicated and/or enforced.</li> <li>Teacher is unaware of what students are doing.</li> <li>Teacher spends too much time dealing with student behavior and disciplinary consequences.</li> <li>Response to student behavior is inconsistent or negative and counterproductive.</li> </ul>
<b>Conveys high expectations, respects students' cultural backgrounds and accommodates diverse needs of students</b>	<ul style="list-style-type: none"> <li>Establishes a culture of excellence that stretches achievement for all students.</li> <li>Student to student interactions are highly respectful and are supportive of one another's individuality and developmental levels.</li> <li>Every student receives verbal and nonverbal indications that he/she is valued and respected.</li> </ul>	<ul style="list-style-type: none"> <li>Encourages a culture in the classroom that challenges all students to continuously improve.</li> <li>Teacher seeks to create an environment where student interactions are respectful.</li> <li>Students receive verbal and nonverbal indications that he/she is valued and respected.</li> </ul>	<ul style="list-style-type: none"> <li>Supports student improvement efforts suitable for most students.</li> <li>Teacher reinforces student to student interactions that are appropriate, but there may be an occasional display of insensitivity to cultural or developmental differences among students.</li> </ul>	<ul style="list-style-type: none"> <li>Reinforces low level learning expectations.</li> <li>Teacher allows displays of insensitivity to cultural or developmental differences among students.</li> </ul>
<b>Models and teaches clear, acceptable communication skills</b>	<ul style="list-style-type: none"> <li>Directions, procedures, and feedback are clear to students and anticipate possible misunderstanding.</li> <li>Teacher's spoken and written language conforms to standard English and contains well chosen vocabulary that enriches the lesson.</li> <li>Teacher finds opportunities to extend student vocabulary.</li> <li>Teacher/student interactions serve as a model of respectful rapport.</li> </ul>	<ul style="list-style-type: none"> <li>Directions, procedures, and feedback are clear to students.</li> <li>Teacher's spoken and written language conforms to standard English.</li> <li>Teacher/student interactions serve as a model of respectful rapport.</li> </ul>	<ul style="list-style-type: none"> <li>Directions, procedures, and feedback are clarified to students after initial student confusion.</li> <li>Teacher's spoken language is audible and when written, is legible.</li> <li>Teacher/student interactions are cordial, but may not transfer to positive student/teacher relations.</li> </ul>	<ul style="list-style-type: none"> <li>Directions, procedures, and feedback are confusing to students.</li> <li>Teacher's spoken language is inaudible to some students.</li> <li>Writing is illegible or contains errors in grammar or syntax.</li> <li>Teacher/student interactions reflect a lack of rapport.</li> </ul>



**Dimension 2: Learning Environment**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Maintains a climate of inquiry</b>	<ul style="list-style-type: none"> <li>Facilitates a collaborative classroom culture that promotes student leadership in learning and in applying a variety of problem solving, inquiry-based strategies.</li> <li>Engages students in problem solving inquiry-based activities through the use of high level questioning techniques, discovery learning, shared inquiry/Socratic discussions.</li> <li>Problem solving culminates in student generated real-world applications.</li> </ul>	<ul style="list-style-type: none"> <li>Engages students in problem solving inquiry-based activities through the use of high level questioning.</li> <li>Student participation and responses indicate individual understanding of content and/or concepts.</li> </ul>	<ul style="list-style-type: none"> <li>Includes some problem solving and inquiry-based activities.</li> <li>Some students are not involved or adequately prepared, do not understand or are not engaged in the processes.</li> <li>A limited number of effective inquiry-based strategies are used.</li> </ul>	<ul style="list-style-type: none"> <li>Instruction does not regularly include problem solving or inquiry-based activities.</li> </ul>
<b>Integrates appropriate technologies</b>	<ul style="list-style-type: none"> <li>Creates a rich learning environment in which students regularly engage in activities that would have been difficult to achieve without technology.</li> <li>Students are actively engaged in using available technology as a tool rather than passively receiving information from the technology.</li> <li>Students engage in ongoing meta-cognitive activities and collaborate with peers and/or experts on educational projects.</li> </ul>	<ul style="list-style-type: none"> <li>Uses technology to construct teaching materials and learning activities.</li> <li>Students are actively engaged in using technology as a tool rather than passively receiving information from the technology.</li> <li>Teaches students to use available computers and other forms of technology to create products.</li> </ul>	<ul style="list-style-type: none"> <li>Demonstrates awareness of and models acceptable use of technology; collaborates with other educational professionals, parents and students, as appropriate.</li> <li>Teacher uses technology to deliver curriculum content to students.</li> <li>Students use technology passively for drill and practice and computer-based training.</li> </ul>	<ul style="list-style-type: none"> <li>Technology to promote organizational efficiency and support instruction is not evident.</li> </ul>

**Dimension 3: Instructional Delivery & Facilitation**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<p><b>Utilizes knowledge of the subject to enrich students' understanding, identifies gaps in students' learning, modifies instruction to respond to student misconceptions</b></p>	<ul style="list-style-type: none"> <li>• Instruction is based on rich content knowledge that is accurate, current and consistent with sound practices of the discipline.</li> <li>• All students are engaged in learning experiences, discussions, questioning, and demonstrations of learning.</li> <li>• Uses a variety of checks for understanding during strategic points in the lesson to identify learning gaps and adjust instruction.</li> <li>• Anticipates problems and uses multiple intervention strategies to assist student understanding and performance.</li> </ul>	<ul style="list-style-type: none"> <li>• Instruction is based on content knowledge that is accurate and current.</li> <li>• Recognizes problems and makes adjustments during the lesson in response to student understanding and performance.</li> <li>• The teacher engages students, models instruction, guides practice and provides feedback, and promotes student demonstration of learning.</li> </ul>	<ul style="list-style-type: none"> <li>• Instruction is based on a general understanding of content knowledge that is accurate but fails to reflect the most current knowledge of the discipline.</li> <li>• Teacher does not check for understanding or recognize gaps in learning during the lesson.</li> <li>• Teacher adjusts future lessons based on student response data at the conclusion of the lesson.</li> </ul>	<ul style="list-style-type: none"> <li>• Instruction references inaccurate, outdated content knowledge.</li> <li>• Makes ineffective use of instructional materials.</li> <li>• Teacher does not use checks for understanding during the lesson.</li> <li>• No adjustments are made to the lesson sequence or pacing when students are confused.</li> </ul>
<p><b>Employs higher order questions</b></p>	<ul style="list-style-type: none"> <li>• Strategic teacher questioning requires students to think critically, problem solve, and defend or justify their answers.</li> <li>• Much of the questioning involves analysis and synthesis.</li> <li>• Students contribute to the discussion by commenting on or adding to answers given by their classmates.</li> <li>• Ideally teachers can take a back seat and discussions can be student-led.</li> </ul>	<ul style="list-style-type: none"> <li>• Most teacher questions encourage thoughtful and extended responses while some are of lower level and can be answered with few words.</li> <li>• Teacher encourages all students to participate.</li> </ul>	<ul style="list-style-type: none"> <li>• Few teacher questioning strategies invite quality responses.</li> <li>• Teacher is quick to provide the correct answers and students are not given the opportunity to think things through.</li> <li>• Only a few students respond.</li> </ul>	<ul style="list-style-type: none"> <li>• Questioning is knowledge level and is solely teacher-driven.</li> <li>• Students appear apathetic and the teacher provides the answers without any effort on their part.</li> </ul>
<p><b>Applies varied instructional strategies and resources</b></p>	<ul style="list-style-type: none"> <li>• Uses multiple resources and an extensive repertoire of instructional strategies that engage and challenge all students and support instructional outcomes.</li> <li>• Teacher stays abreast of current research through professional organizations, universities, local resources, and through the use of technology.</li> <li>• Teacher incorporates innovative ideas and materials.</li> </ul>	<ul style="list-style-type: none"> <li>• Shows awareness of resources for students available through the school and district and some familiarity with resources external to the school.</li> <li>• Teacher uses multiple instructional strategies to maintain focus, engage students and support instructional outcomes for all students.</li> </ul>	<ul style="list-style-type: none"> <li>• Makes use of a limited number of instructional strategies to support student understanding.</li> <li>• Teacher is aware of and uses resources available through the school and district but does not use resources available more broadly.</li> </ul>	<ul style="list-style-type: none"> <li>• Develops lessons that lack instructional focus and do not engage students and support instructional outcomes.</li> <li>• Teacher does not use resources available through the school and district to enhance content.</li> <li>• Classroom time is often filled with activities that merely consume time and do not contribute to instructional outcomes.</li> </ul>

**Dimension 3: Instructional Delivery & Facilitation**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Delivers engaging, challenging and relevant lessons</b>	<ul style="list-style-type: none"> <li>Lessons (<i>and units</i>) are designed to maximize productive time.</li> <li>High levels of rigor and relevance challenge students to be intellectually engaged throughout.</li> <li>Teacher clearly demonstrates and articulates how content relates and applies to instructional activities, life, work and community.</li> <li>Students can explain how the lesson relates to other disciplines and to themselves.</li> </ul>	<ul style="list-style-type: none"> <li>Lessons (<i>and units</i>) are designed to provide students with rigorous curriculum that is intellectually engaging.</li> <li>Lessons often demonstrate the subject's importance and relation to life situations, and students can state how the subject being studied is relevant to their own learning.</li> </ul>	<ul style="list-style-type: none"> <li>Lessons (<i>and units</i>) spend the majority of the class time on knowledge level tasks that may require only some intellectual engagement.</li> <li>Lessons do not consistently demonstrate an understanding of the subject's relationship to life situations.</li> <li>Whether or not they find it interesting, the students follow the teacher's directions and do what is asked of them.</li> </ul>	<ul style="list-style-type: none"> <li>Lessons (<i>and units</i>) are too easy for the majority of students and class time is spent on busy work with low intellectual engagement.</li> <li>Lessons do not help the students make the connection between the instructional goals and what they are learning.</li> <li>Presentation of the lesson is not interesting and the students fail to apply the lesson to themselves.</li> </ul>
<b>Differentiates instruction</b>	<ul style="list-style-type: none"> <li>Differentiates instruction efficiently and appropriately by content, process and product to address the unique learning differences of students.</li> <li>Both daily instruction and unit design clearly address the wide range of learning styles and abilities present in the classroom.</li> <li>Teacher provides opportunities for student choice in demonstrating mastery of learning goals.</li> </ul>	<ul style="list-style-type: none"> <li>Provides evidence of incorporating various differentiated instructional strategies (<i>e.g. ability grouping or compacting of lessons</i>) to meet the needs of students with varying learning styles and abilities.</li> </ul>	<ul style="list-style-type: none"> <li>Provides little evidence of incorporating various differentiated instruction.</li> <li>Recognizes the need for differentiation but requires support to differentiate instruction for students with varying learning styles and abilities.</li> </ul>	<ul style="list-style-type: none"> <li>Does not differentiate instruction to address the needs of students with varying learning styles and abilities.</li> </ul>
<b>Provides immediate and specific feedback to students</b>	<ul style="list-style-type: none"> <li>Provides timely, deliberate, and consistent feedback during guided practice, discussion and major activities.</li> <li>Students use feedback in their learning to improve their performance.</li> </ul>	<ul style="list-style-type: none"> <li>Provides timely and consistent feedback.</li> <li>Students are clear on what areas need improvement for the next assigned task.</li> </ul>	<ul style="list-style-type: none"> <li>Provides feedback at the end of the unit.</li> <li>Feedback does not provide specific information for students to make error corrections or is not timely enough to improve performance.</li> </ul>	<ul style="list-style-type: none"> <li>Provides inadequate or no feedback: feedback provided is not provided in a timely manner.</li> </ul>

### Dimension 4: Assessment

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<p><b>Analyzes and applies data from multiple measures to diagnose students' learning needs, inform instruction, and monitor progress.</b></p>	<ul style="list-style-type: none"> <li>• Consciously plans and uses pre-assessments, formative and summative assessments in a continuous improvement cycle based on student achievement.</li> <li>• All assessment criteria and standards are clear to students and contribute to their learning.</li> <li>• Students and the teacher analyze data results from multiple assessments to make decisions about their progress and develop appropriate interventions relative to their needs.</li> <li>• Progress monitoring data is regularly used to provide specific feedback to individual students and analyze teaching results.</li> <li>• Students make changes in their learning paths based on the analysis of assessment data.</li> </ul>	<ul style="list-style-type: none"> <li>• Uses pre-assessment, formative assessment, and summative assessment data strategically to set and adjust learning goals based on student needs.</li> <li>• All assessment criteria and standards are clear.</li> <li>• Gathers data during instruction and uses data to inform instruction.</li> <li>• Analyzes multiple assessment data regularly and adjusts instruction based on analysis.</li> <li>• Assessment methodologies may be differentiated for groups of students.</li> </ul>	<ul style="list-style-type: none"> <li>• Regularly uses summative assessment data to determine that students have achieved instructional outcomes.</li> <li>• Assessment criteria and standards have been developed, but may not be clear.</li> <li>• Use of formative assessment data is rudimentary, assessing some instructional outcomes.</li> <li>• Little intervention or differentiation is evident.</li> </ul>	<ul style="list-style-type: none"> <li>• Assessment procedures are not congruent with instructional outcomes.</li> <li>• No criteria or standards for assessment are evident.</li> <li>• Teacher does not incorporate assessment data to monitor and adjust instruction or to differentiate assessment for individuals.</li> <li>• Assessment measures are not used to diagnose student learning or to develop interventions to improve instruction.</li> </ul>
<p><b>Designs and uses formative and summative assessments that lead to mastery</b></p>	<ul style="list-style-type: none"> <li>• Formative and summative assessments are developed as a part of unit design prior to instruction.</li> <li>• Performance results are used to build a shared understanding with students of what they should know and be able to do as a result of instruction.</li> <li>• Formative assessment tasks are embedded within the lesson and/or require a performance component for students to demonstrate what they know and are able to do.</li> <li>• The teacher evaluates the results from the formative assessment tasks and uses that information to re-teach or improve future lessons.</li> </ul>	<ul style="list-style-type: none"> <li>• Some formative assessments are used to measure achievement in both content and higher-level thinking.</li> <li>• Summative assessments are given at the end of lessons or units and provide varied options for students to demonstrate what they know and are able to do.</li> <li>• Results of the assessments are the basis for instructional planning and used to evaluate overall student progress toward meeting the standards.</li> </ul>	<ul style="list-style-type: none"> <li>• Formative and summative assessments are developed but may not measure beyond understanding of content or may not use the tools to help students understand individual achievement.</li> <li>• Assessments after instruction are not used for instructional planning but evaluate student progress toward meeting standards.</li> <li>• Some summative assessment content material is considered prior to instructional lesson and unit design.</li> </ul>	<ul style="list-style-type: none"> <li>• No formative assessment is given.</li> <li>• A single type of classroom assessment is used that may or may not be aligned with the curriculum and is used only to justify student grades.</li> <li>• Lessons and units are designed prior to assessments being administered.</li> </ul>

### Dimension 4: Assessment

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Modifies teacher made assessments to accommodate diversity</b>	<ul style="list-style-type: none"> <li>• Efficiently and appropriately differentiates assessments by content, process and/or product to address the unique learning differences of students that have a wide range of learning styles and abilities.</li> <li>• Students are aware of their learning styles and given choices to demonstrate their learning.</li> </ul>	<ul style="list-style-type: none"> <li>• Appropriately differentiates assessments to address the unique learning differences of students that have a wide range of learning styles and abilities.</li> <li>• Provides a variety of assessments to meet the needs of students.</li> <li>• Students may have choices in their assessments.</li> </ul>	<ul style="list-style-type: none"> <li>• Little evidence is provided that assessments are differentiated to meet the needs of students' learning styles or abilities.</li> </ul>	<ul style="list-style-type: none"> <li>• Assessments are "one size fits all."</li> </ul>
<b>Communicates assessment data to students and parents</b>	<ul style="list-style-type: none"> <li>• Assessment criteria, due dates, and grading methodology are clearly communicated in Edline in such a way that promotes student learning.</li> <li>• Teacher is proactive in creating and maintaining frequent and effective two-way communication with students and parents.</li> <li>• Grades and progress reports are completed and updated regularly to allow students to monitor their own proficiency.</li> </ul>	<ul style="list-style-type: none"> <li>• Grades and progress reports are completed and updated regularly in Edline to allow students to check their own progress.</li> <li>• Teacher has designed a system for documenting and reporting student learning. This is shared with students and parents in advance.</li> <li>• Teachers lead conferences with parents and students to share assessment data with parents.</li> </ul>	<ul style="list-style-type: none"> <li>• Grading and reporting of student learning is completed at the end of the unit.</li> <li>• Teacher does not consistently communicate progress with students and parents in a timely fashion.</li> <li>• Grades and progress reports are completed per the school schedule.</li> </ul>	<ul style="list-style-type: none"> <li>• Grading is inconsistent and may be subjective.</li> <li>• Students and parents do not know what the cumulative grade will be until the official report comes out.</li> <li>• Grades and progress reports are not completed per the school schedule.</li> </ul>

### Dimension 5: Professional Responsibilities & Ethical Conduct

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Demonstrates punctuality, attendance, and timely completion of records and reports</b>	<ul style="list-style-type: none"> <li>• Complies fully with all school rules, policies and procedures regarding punctuality and attendance.</li> <li>• Has developed a system for maintaining information on students, lesson plans, reports and other data which is efficient, current and useful.</li> </ul>	<ul style="list-style-type: none"> <li>• Complies with school rules, policies and procedures.</li> <li>• The system for maintaining information on students, lesson plans, reports and other data is effective.</li> </ul>	<ul style="list-style-type: none"> <li>• Inconsistently complies with school rules, policies and procedures.</li> <li>• The system for maintaining student information, lesson plans, reports and other data is rudimentary and only partially effective.</li> </ul>	<ul style="list-style-type: none"> <li>• Does not comply with school rules, policies and procedures.</li> <li>• There is no system of record keeping, lesson plans, records or other data or the system is in disarray.</li> </ul>
<b>Performs assigned duties and complies with policies, procedures, programs, and the Code of Ethics</b>	<ul style="list-style-type: none"> <li>• Fully understands, abides by and supports the legal and professional responsibilities pertaining to education.</li> <li>• The teacher complies with district and school policies, rules, procedures and the Code of Ethics.</li> <li>• The teacher acts proactively in fostering compliance amongst colleagues with standards of excellence and demonstrates leadership in exhibiting a high ethical standard of decorum throughout the school setting.</li> </ul>	<ul style="list-style-type: none"> <li>• Understands and abides by the legal and professional responsibilities pertaining to education.</li> <li>• The teacher complies fully with district and school policies, rules, procedures and the Code of Ethics.</li> <li>• The teacher models professional and ethical standards of decorum throughout the school setting.</li> </ul>	<ul style="list-style-type: none"> <li>• Demonstrates a limited understanding through inconsistent behaviors.</li> <li>• Inconsistently complies with district and school policies, rules, procedures and/or the Code of Ethics.</li> <li>• The teacher exhibits minimal deference to professional and ethical standards of decorum throughout the school setting.</li> </ul>	<ul style="list-style-type: none"> <li>• Disregards or has no awareness of legal and professional responsibilities pertaining to education.</li> <li>• The teacher does not comply with district and school policies, rules, procedures and/or the Code of Ethics.</li> <li>• The teacher's behavior is below acceptable ethical standards of decorum throughout the school setting.</li> </ul>
<b>Demonstrates professionalism</b>	<ul style="list-style-type: none"> <li>• Displays conduct based on the highest professional standards.</li> <li>• Acts honestly and with integrity when dealing with students, parents, colleagues and/or the community.</li> <li>• The teacher demonstrates a pattern of participation in district and school initiatives, contributes to decision-making processes, serves on sub-committees and/or disseminates information when appropriate.</li> </ul>	<ul style="list-style-type: none"> <li>• Deals with students, parents and/or colleagues with honesty and integrity.</li> <li>• The teacher participates in and implements school and district initiatives.</li> <li>• The teacher clearly implements the evidence.</li> </ul>	<ul style="list-style-type: none"> <li>• Uses poor judgment when dealing with students, parents and/or colleagues.</li> <li>• The teacher implements most decisions made at the school and district levels.</li> </ul>	<ul style="list-style-type: none"> <li>• Displays unethical or dishonest conduct when dealing with students, parents and/or colleagues.</li> <li>• The teacher does not implement decisions made at school or district levels.</li> </ul>

**Dimension 6: Relationship with Students**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Demonstrates Knowledge of Students</b>	<ul style="list-style-type: none"> <li>Teacher recognizes the value of understanding students' interests and cultural heritage and displays this knowledge by using interest surveys, invitational relationship-building activities, and equitable procedures.</li> <li>Teacher displays understanding of individual students' skill, knowledge, and language proficiency and has a strategy for maintaining such information.</li> <li>Teacher possesses information about each student's learning and medical needs, collecting such information from a variety of sources.</li> <li>Teacher displays extensive understanding of how students learn and applies this knowledge to develop a positive relationship with individual students.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher recognizes the value of understanding students' interests and cultural heritage and displays this knowledge for groups of students.</li> <li>Teacher recognizes the value of understanding students' skills, knowledge, and language proficiency and displays this knowledge for groups of students.</li> <li>Teacher is aware of students' special learning and medical needs.</li> <li>Teacher's knowledge of how students learn is accurate and current.</li> <li>Teacher applies this knowledge in a positive relationship to the class as a whole and to groups of students.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher recognizes the value of understanding students' interests and cultural heritage but displays this knowledge only to the class as a whole.</li> <li>Teacher recognizes the value of understanding students' skills, knowledge, and language proficiency but displays this knowledge only for the class as a whole.</li> <li>Teacher displays awareness of the importance of knowing students' special learning or medical needs, but such knowledge may be incomplete or inaccurate.</li> <li>Teacher recognizes the value of knowing how students learn, but this knowledge is limited or outdated.</li> <li>Students appear to lack a positive relationship with the teacher.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher displays little or no knowledge of students' interests or cultural heritage and does not indicate that such knowledge is valuable.</li> <li>Teacher displays little or no knowledge of students' skills, knowledge, and language proficiency and does not indicate that such knowledge is valuable.</li> <li>Teacher displays little or no understanding of students' special learning or medical needs or why such knowledge is important to a positive relationship.</li> <li>Teacher does not seek information about how students learn and appears to lack positive relationships with them.</li> </ul>
<b>Builds Relationships Through Instructional Interactions</b>	<ul style="list-style-type: none"> <li>Teacher allows and encourages all students to be part of class discussions and interactions.</li> <li>Teacher brings student interests into the content.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher allows and encourages most students to be part of class discussions and interactions.</li> <li>Teacher brings some student interests into the content.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher allows some students to be involved in class discussions and interactions.</li> <li>There is little evidence of the teacher's bringing student interests into the content.</li> <li>Teacher does not differentiate content, process, and/or assessment.</li> <li>Teacher provides little or no feedback on the students' work.</li> <li>Teacher does not provide feedback to individual students.</li> <li>Teacher does not convey high expectations to the class.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher does not engage students in class discussions and interactions.</li> <li>Awareness of student interests is not evident in content delivery.</li> </ul>

**Dimension 6: Relationship with Students**

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<p><b>Creates a Positive Environment of Respect and Rapport</b></p>	<ul style="list-style-type: none"> <li>• Teacher interactions with students reflect genuine respect and caring for individuals as well as groups of students.</li> <li>• Students appear to trust the teacher with sensitive information.</li> <li>• Students demonstrate genuine caring for one another and monitor one another's treatment of peers, correcting classmates respectfully when needed.</li> </ul>	<ul style="list-style-type: none"> <li>• Teacher interactions with students reflect the same amount of respect given to and expected from every student.</li> <li>• Teacher-student interactions are friendly and demonstrate general caring and respect. Such interactions are appropriate to the age and cultures of the students.</li> </ul>	<ul style="list-style-type: none"> <li>• Teacher-student interactions are generally appropriate but may reflect occasional inconsistencies, favoritism, or disregard for students' cultures.</li> <li>• Students exhibit only minimal respect for the teacher.</li> <li>• Students are disrespectful to one another.</li> </ul>	<ul style="list-style-type: none"> <li>• Teacher interaction with at least some students is negative, demeaning, sarcastic, or inappropriate to the age or culture of the students.</li> <li>• Students exhibit disrespect for the teacher.</li> <li>• Student interactions are characterized by conflict, sarcasm, or put-downs.</li> </ul>



### Dimension 7: Relationship with Parents and Community: Building Relationships with Parents

Elements	Distinguished (3 pts)	Proficient (2 pts)	Professional Support Needed (1 pt)	Unsatisfactory (0 pts)
<b>Establish contact with parents early in the school year</b>	<ul style="list-style-type: none"> <li>Use of multiple methods to reach out and welcome all parents within the first two weeks of the school year and elicits feedback from parents about their child.</li> </ul>	<ul style="list-style-type: none"> <li>Welcome parents with a personal contact early in the school year (<i>phone calls, welcome letter, home visits</i>).</li> </ul>	<ul style="list-style-type: none"> <li>Attempts to contact parents early in the school year, but may not have a consistent plan and/or needs guidance to do so effectively.</li> </ul>	<ul style="list-style-type: none"> <li>Makes no effort to contact parents early in the school year.</li> </ul>
<b>Establish a relationship of mutual trust and respect with parents</b>	<ul style="list-style-type: none"> <li>Consistently considers family background and cultural diversity when working with parents.</li> <li>Demonstrates sensitivity, respect, and understanding of families to build positive relationships with all families.</li> </ul>	<ul style="list-style-type: none"> <li>Demonstrates awareness of family background and cultural diversity.</li> <li>Demonstrates sensitivity, respect, and understanding of families to build a positive relationship.</li> </ul>	<ul style="list-style-type: none"> <li>Lacks awareness of cultural diversity of families.</li> <li>Needs guidance to learn effective strategies for building positive relationships with parents.</li> <li>Does not establish a relationship of trust and mutual respect with parents.</li> </ul>	<ul style="list-style-type: none"> <li>Does not maintain a family friendly learning environment.</li> <li>Does not consistently reach out to invite parents to collaborate and participate in learning in innovative ways.</li> </ul>
<b>Maintain a family friendly learning environment</b>	<ul style="list-style-type: none"> <li>Consistently reaches out to invite parents to collaborate and participate in learning in innovative ways.</li> </ul>	<ul style="list-style-type: none"> <li>Maintains a family friendly environment that encourages engagement (<i>Inviting climate, opportunities to volunteer in the classroom, and encourage collaboration with parents</i>).</li> </ul>	<ul style="list-style-type: none"> <li>Attempts to maintain a family friendly environment that encourages engagement, but needs guidance to do so effectively.</li> </ul>	<ul style="list-style-type: none"> <li>Does not maintain a family friendly environment that encourages engagement.</li> </ul>
<b>Maintains two-way communication with parents</b>	<ul style="list-style-type: none"> <li>Multiple methods of two-way communication are used regularly to maintain communication with all parents (<i>See examples in proficient</i>).</li> <li>Consistently implements strategies to maximize participation of all parents in parent teacher conferences (<i>See examples in proficient</i>).</li> <li>Student-led conferences are evident when appropriate.</li> </ul>	<ul style="list-style-type: none"> <li>Two-way communication is used to keep parents informed (<i>agendas, Edline, email, phone calls</i>).</li> <li>Implements strategies to maximize participation in parent-teacher conferences (<i>offering multiple locations, convenient times, opportunities for student-led conferences, follow up with parents who do not initially reply</i>).</li> </ul>	<ul style="list-style-type: none"> <li>Inconsistent in communication with parents.</li> <li>May need strategies to implement two-way communication.</li> <li>Minimal effort to encourage parent participation in parent-teacher conferences.</li> <li>Does not implement multiple efforts to reach parents and may only conferences with parents when there is a problem.</li> </ul>	<ul style="list-style-type: none"> <li>Does not maintain communication with parents.</li> <li>Makes no effort to encourage parent participation in parent-teacher conferences.</li> </ul>
<b>Responsive to parent needs</b>	<ul style="list-style-type: none"> <li>Consistently accessible to parents, actively listens to concerns, immediately addresses issues, elicits feedback regularly, and engages in open dialogue.</li> </ul>	<ul style="list-style-type: none"> <li>Accessible to parents, listens to concerns, addresses issues in a timely manner and engages in open dialogue.</li> </ul>	<ul style="list-style-type: none"> <li>Limited accessibility to parents and does not consistently respond to parents in a timely manner.</li> </ul>	<ul style="list-style-type: none"> <li>Not accessible to parents.</li> <li>Not responsive to parent concerns.</li> </ul>

**Dimension 7: Relationship with Parents and Community: Engaging Parents in Learning**

<b>Elements</b>	<b>Distinguished (3 pts)</b>	<b>Proficient (2 pts)</b>	<b>Professional Support Needed (1 pt)</b>	<b>Unsatisfactory (0 pts)</b>
<b>Promotes parent understanding of academic standards and expectations</b>	<ul style="list-style-type: none"> <li>Provides all parents with information, specific to their child, about standards and expectations in a format that parents can understand.</li> </ul>	<ul style="list-style-type: none"> <li>Provides parents with information about academic standards and expectations in a format that parents can understand (<i>Open House, handouts, post on website, parent conferences</i>).</li> </ul>	<ul style="list-style-type: none"> <li>Limited information provided to parents about academic standards and expectations.</li> <li>May need assistance in providing it in a format that parents can understand.</li> </ul>	<ul style="list-style-type: none"> <li>Does not provide parents with information about academic standards and expectations.</li> </ul>
<b>Reports to parents about student performance and academic progress</b>	<ul style="list-style-type: none"> <li>Consistently interprets and explains individual students' academic progress (<i>assessment results, accomplishments, and needs for improvement</i>) in a manner that provides parents with an ongoing picture of children's performance.</li> </ul>	<ul style="list-style-type: none"> <li>Interprets and explains students' academic progress (<i>assessment results, accomplishments, and needs for improvement</i>) in a manner that provides parents with an accurate picture of children's performance.</li> </ul>	<ul style="list-style-type: none"> <li>Limited information shared with parents on student progress.</li> <li>Does not provide an accurate picture of children's performance in a timely manner.</li> </ul>	<ul style="list-style-type: none"> <li>Does not provide parents with information about students' performance and academic progress.</li> </ul>
<b>Provides opportunities to assist with student learning at home</b>	<ul style="list-style-type: none"> <li>Consistently provides all parents with multiple strategies and opportunities to assist with student learning (See examples for proficient).</li> <li>Offers individualized support.</li> </ul>	<ul style="list-style-type: none"> <li>Provides strategies and opportunities for parents to assist with student learning at home (parent workshops, meetings, suggestions for home activities, access to resources).</li> </ul>	<ul style="list-style-type: none"> <li>Provides limited strategies and opportunities for parents to assist with student learning.</li> </ul>	<ul style="list-style-type: none"> <li>Does not provide or promote opportunities for parents to assist with student learning at home.</li> </ul>
<b>Provides information about school and community resources to parents</b>	<ul style="list-style-type: none"> <li>Consistently provides school and community event information in newsletters, email, or Edline about available resources.</li> <li>Consistently provides information to parents about community events and resources.</li> </ul>	<ul style="list-style-type: none"> <li>Provides school and community event information in newsletters, email, or Edline about available resources.</li> <li>Provides timely information to parents about community events and resources.</li> </ul>	<ul style="list-style-type: none"> <li>Limited information is provided to parents in newsletters, email, or Edline about available resources.</li> <li>Limited information is provided to parents about community events and resources.</li> </ul>	<ul style="list-style-type: none"> <li>Does not provide information to parents in newsletters, email, or Edline about available resources.</li> <li>Does not provide information to parents about community events and resources.</li> </ul>
<b>Provides meaningful connections with learning and community</b>	<ul style="list-style-type: none"> <li>Consistently integrates everyday activities to real-life situations.</li> <li>Consistently integrates community service opportunities with enhanced learning.</li> </ul>	<ul style="list-style-type: none"> <li>Integrates everyday activities to real-life situations.</li> <li>Integrates community service opportunities with learning.</li> </ul>	<ul style="list-style-type: none"> <li>Limited integration of everyday activities to real-life situations.</li> <li>Limited integration of community service opportunities with learning.</li> </ul>	<ul style="list-style-type: none"> <li>Does not integrate everyday activities to real-life situations.</li> <li>Does not integrate community service opportunities with learning.</li> </ul>

APPENDIX J  
BPS INSTRUCTIONAL PERSONNEL EVALUATION INSTRUMENT



BPS Instructional Personnel Evaluation Instrument

Name \_\_\_\_\_ School Year \_\_\_\_\_
Last First MI
School/Dept Name \_\_\_\_\_ School/Dept # \_\_\_\_\_ Contract Status: INTERIM [ ] ANNUAL [ ]
Principal/Dept Head \_\_\_\_\_ Assignment \_\_\_\_\_ New Hire [ ] AC [ ] CC/PSC [ ]

PERFORMANCE AREAS

RATINGS

A. Quality of Instruction (Must be completed for all certificated employees):

Instructional Design and Lesson Planning

- 1. Sets instructional outcomes and aligns instruction with state-adopted standards.
2. Designs instruction using student prior knowledge and diagnostic student data to build coherent lessons.
3. Designs ways to monitor learning, designs use of appropriate instructional resources, and designs formative assessments.
4. Requires students to understand and demonstrate skills and competencies.

Rating scale table with columns: Distinguished (3), Proficient (2), Professional Support Needed (1), Unsatisfactory (0). Rows: SELF, ADMIN.

SELF [ ] ADMIN [ ]

Teacher Comments:
Administrator Comments:

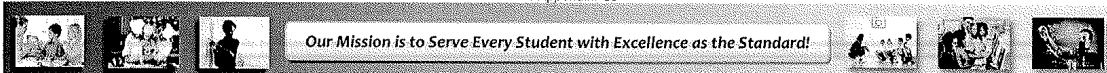
Learning Environment

- 1. Creates and maintains a safe and organized learning environment.
2. Promotes a flexible, inclusive, collaborative, and student-centered learning environment.
3. Allocates and manages time, space, resources.
4. Manages student conduct.
5. Conveys high expectations, respects students' cultural backgrounds and accommodates diverse needs of students.
6. Models and teaches clear, acceptable communication skills.
7. Maintains a climate of inquiry.
8. Integrates appropriate technologies.

Rating scale table with columns: Distinguished (3), Proficient (2), Professional Support Needed (1), Unsatisfactory (0). Rows: SELF, ADMIN.

SELF [ ] ADMIN [ ]

Teacher Comments:
Administrator Comments:





### Instructional Delivery and Facilitation

- Utilizes knowledge of the subject to enrich students' understanding, identifies gaps in students' learning, and modifies instruction to respond to student misconceptions.
- Employs higher order questions.
- Applies varied instructional strategies and resources.
- Delivers engaging, challenging and relevant lessons.
- Differentiates instruction.
- Provides immediate and specific feedback to students.

Distinguished		Proficient		Professional Support Needed		Unsatisfactory	
3		2		1		0	
SELF	ADMIN	SELF	ADMIN	SELF	ADMIN	SELF	ADMIN

SELF

ADMIN

Teacher Comments:

---

Administrator Comments:

---

### Assessment

- Analyzes and applies data from multiple measures to diagnose students' learning needs, inform instruction, and monitor progress.
- Designs and uses formative and summative assessments that lead to mastery.
- Modifies teacher made assessments to accommodate diversity.
- Communicates assessment data to students and parents.

Distinguished		Proficient		Professional Support Needed		Unsatisfactory	
3		2		1		0	
SELF	ADMIN	SELF	ADMIN	SELF	ADMIN	SELF	ADMIN

SELF

ADMIN

Teacher Comments:

---

Administrator Comments:

---

### B. Quality of Instruction *(Must be completed for all certificated employees):*

#### Professional Responsibility and Ethical Conduct

- Demonstrates punctuality, attendance, and timely completion of records and reports.
- Performs assigned duties and complies with policies, procedures, programs, and the Code of Ethics.
- Demonstrates professionalism.
- Initiates professional communication with appropriate stakeholders.
- Applies technology to organize and communicate assessment information.

Distinguished		Proficient		Professional Support Needed		Unsatisfactory	
3		2		1		0	
SELF	ADMIN	SELF	ADMIN	SELF	ADMIN	SELF	ADMIN

SELF

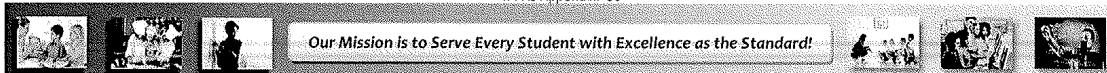
ADMIN

Teacher Comments:

---

Administrator Comments:

---





### Relationships with Students

1. Displays knowledge and understanding of how students learn and applies knowledge to building positive relationships with students.
2. Teacher brings student interest into content.
3. Teacher interactions with students reflect respect and caring for the individual as well as groups of students.

Distinguished		Proficient		Professional Support Needed		Unsatisfactory	
3		2		1		0	
SELF	ADMIN	SELF	ADMIN	SELF	ADMIN	SELF	ADMIN

SELF

ADMIN

Teacher Comments:

---

Administrator Comments:

---

### Relationships with Parents and Community

1. Establishes early contact with parents.
2. Establishes a relationship of mutual trust and respect with parents.
3. Maintains a family-friendly learning environment.
4. Maintains two-way communication with parents.
5. Responsive to parent needs.
6. Promotes parent understanding of academic standards and expectations.
7. Reports to parents about student performance and progress.
8. Provides opportunities to assist student learning at home.
9. Provides information about school and community resources to parents.
10. Provides meaningful connections with learning and the community.

Distinguished		Proficient		Professional Support Needed		Unsatisfactory	
3		2		1		0	
SELF	ADMIN	SELF	ADMIN	SELF	ADMIN	SELF	ADMIN

SELF

ADMIN

Teacher Comments:

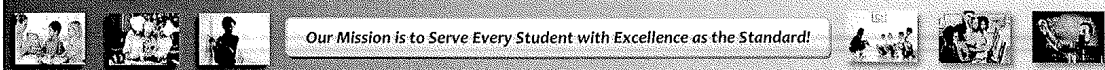
---

Administrator Comments:

---

#### A. FORMATIVE SCORING BY ADMINISTRATOR:

Component	Possible Points	Points Earned
Total Observations Points	21	





**Principal Comments:**

---

---

---

**Teacher Comments (Optional):**

---

---

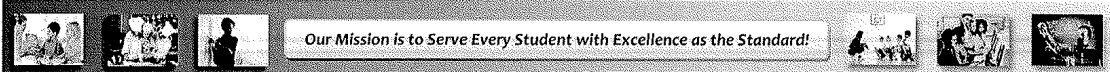
---

**My signature indicates that this evaluation has been discussed with me:**

\_\_\_\_\_/\_\_\_\_\_  
Signature of Employee (Blue Ink Only) Date

\_\_\_\_\_/\_\_\_\_\_  
Signature of Principal/Administrative Supervisor (Required) (Blue Ink Only) Date

\_\_\_\_\_/\_\_\_\_\_  
Signature of Assistant Principal (Required) (Blue Ink Only) Date



APPENDIX K  
INSTRUCTIONAL PROFESSIONAL GROWTH PLAN TEMPLATE





**Teacher Professional Growth Plan (PGP) Goal**

Teacher Number \_\_\_\_\_ School \_\_\_\_\_

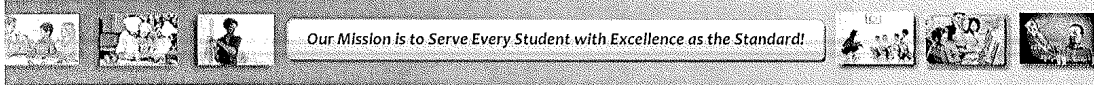
RATIONALE FOR PGP GOAL:			
Data Sources	Best Practice	Reflection	"Stretch"
<i>Multiple sources Quantitative - student &amp; professional practice Qualitative - student &amp; professional practice</i>	<i>Connected to data analysis Research within school Research beyond school</i>	<i>Connected to data analysis Comparison: self-reflection and best-practice research school Link improved practice to student needs</i>	<i>Connected to data analysis Linked to SIP Outcome: new/ strengthened practice</i>

**PROFESSIONAL GROWTH PLAN GOAL:**

**WORK PLAN STRATEGIES:** *connected to goal; action-oriented; sustainable; defined in-process measures; planned feedback & reflection; "doable" timelines; defined new or improved practice*

#	Strategy	Grading Period Implementation				Specific timelines	In-process measures
		1	2	3	4		

**OUTCOME MEASURES AND REFLECTION:** *quantitative & qualitative student and professional practice outcomes*



APPENDIX L  
INDIVIDUAL PROFESSIONAL GROWTH PLAN DEVELOPMENT  
RUBRICS

## Individual Professional Growth Plan (PGP) *Development*

**Plan Development** (*based on continuous improvement cycle*) – this section is to be completed by the school principal (or designee) and three teacher-leaders upon thorough review of each teacher’s annual Professional Growth Plan. The three teacher-leader assessments will be added together and then averaged for a sub-final score. The administrator assessment will be added to the teacher-leader sub-final score and then divided by two. The final PGP points will then be rounded to the nearest tenth of a decimal.

Elements	Distinguished (4 pts)	Proficient (3 pts)	Professional Support Needed (2 pts)	Unsatisfactory (0 pts)
<b>Development of Professional Growth Plan Goal</b>	<p>Uses a researched, data-informed rationale to develop and analyze goals.</p> <ul style="list-style-type: none"> <li>● Seeks out and accurately draws inferences from multiple data sources including the future learning management system.</li> <li>● Uses both quantitative and qualitative student performance data.</li> <li>● Teacher articulates research based rationale for his/her goal including best practice guidance from within and beyond the school.</li> <li>● Shows evidence of means to inform and involve students in data analysis and instructional delivery improvement efforts.</li> <li>● Analyzes student assessment data to develop goal/goals that are linked to classroom practice and the school improvement plan.</li> <li>● Has supported evidence to show teacher reviewed his/her professional skills and can show how the objectives connect to both strengthened instructional practice and improved student achievement outcome.</li> <li>● Goal is clearly identified "stretch" for teacher based on current classroom practice.</li> </ul>	<p>Uses a researched, data-informed rationale to develop his/her goal.</p> <ul style="list-style-type: none"> <li>● Makes accurate use of data.</li> <li>● Uses quantitative or qualitative student performance data.</li> <li>● Teacher articulates research based rationale for his/her goal including best practice guidance from within the school.</li> <li>● Shows evidence of means to inform and involve students in data analysis or instructional delivery improvement efforts.</li> <li>● Analyzes student assessment data to develop goal/goals that are linked to classroom practice and the school improvement plan.</li> <li>● Has supported evidence to show teacher reviewed his/her professional skills and can show how the objectives connect to both strengthened professional practice and improved student achievement outcome.</li> <li>● Goal is somewhat of a "stretch" based on current classroom practice.</li> </ul>	<ul style="list-style-type: none"> <li>● Uses student achievement data as required to design individual goals.</li> <li>● May have evidence of use of only quantitative student performance data.</li> <li>● Links to school improvement plan is limited.</li> <li>● Little evidence of reflection of current practice.</li> <li>● Goal is not a "stretch" goal based on current classroom practice.</li> </ul>	<ul style="list-style-type: none"> <li>● Has difficulty providing evidence of using student achievement data to design goal.</li> <li>● No evidence of link to school improvement plan.</li> <li>● No evidence of reflection of current practice.</li> <li>● Sets low performance goal.</li> </ul>

### Individual Professional Growth Plan (PGP) *Development*

Elements	Distinguished (4 pts)	Proficient (3 pts)	Professional Support Needed (2 pts)	Unsatisfactory (0 pts)
<b>Work Plan Strategies</b>	<p>There are defined qualitative and quantitative in-process measuring elements related to refined instructional practice and enhanced student mastery.</p> <ul style="list-style-type: none"> <li>● There is evidence of defined learning strategies and professional development to influence changes in the teacher's instructional practices.</li> <li>● Strategies are action oriented and sustainable.</li> <li>● There is a clear, direct connection to the professional growth goal.</li> <li>● Timelines are defined and realistic.</li> <li>● The strategies clearly define elements of new or improved professional practice that aligns with B.E.S.T.</li> <li>● New practice is followed by feedback and/or reflection.</li> <li>● Feedback is planned from both peers and administration.</li> </ul>	<p>There are defined qualitative or quantitative in-process measuring elements related to refined instructional practice and enhanced student mastery.</p> <ul style="list-style-type: none"> <li>● There is evidence of defined learning strategies and professional development to influence changes in the teacher's instructional practices.</li> <li>● Strategies are action oriented and sustainable.</li> <li>● Connection to the professional growth goal is evident, but weak.</li> <li>● Timelines are somewhat defined and/or not realistic.</li> <li>● The strategies address elements of new or improved professional practice that aligns with B.E.S.T.</li> <li>● Strategies for feedback and/or reflection of new practices are not clearly defined, or lack multiple feedback opportunities for formative assessment of new practice.</li> </ul>	<p>There are qualitative or quantitative in-process monitoring elements that are weakly related to improved instructional practice and improved student mastery.</p> <ul style="list-style-type: none"> <li>● Limited evidence of defined learning strategies and/or professional development to influence changes in the teacher's instructional practices.</li> <li>● Strategies may be action oriented, but lack sustainability.</li> <li>● Connection to the professional growth goal is weak.</li> <li>● Timelines are not clearly defined and/or realistic.</li> <li>● Strategies for feedback and/or reflection or new practices are limited.</li> </ul>	<ul style="list-style-type: none"> <li>● No evidence of defined learning strategies and/or professional development to influence changes in the teacher's instructional practices.</li> <li>● Strategies are not action oriented and/or lack sustainability.</li> <li>● No connection to the professional growth goal.</li> <li>● Timelines are poorly defined and/or unrealistic.</li> <li>● In process measuring elements are poorly defined with limited or no linkage to refined instructional practices or improved student mastery.</li> <li>● No strategies for feedback and/or reflection of new practices.</li> </ul>
	<b>Distinguished (2 pts)</b>	<b>Proficient (1 pt)</b>	<b>Professional Support Needed (0 pts)</b>	<b>Unsatisfactory (0 pts)</b>
<p><b>Outcome Measures and Reflections</b></p> <p>Connected to data; identified need; Strategic and Specific, Measurable, Attainable, Results-oriented and time bound (SMART)</p>	<ul style="list-style-type: none"> <li>● The PGP includes target goals that are explicitly connected to quantitative and qualitative data from above ("Data Informed") and are directly based on changes in professional practice.</li> <li>● Specific student outcomes and defined, improved practice are clearly identified.</li> </ul>	<ul style="list-style-type: none"> <li>● The PGP target goals are connected to quantitative and/or qualitative measurable data and can be explained by changes in professional practice.</li> <li>● Student outcomes and improved practice are identified.</li> </ul>	<ul style="list-style-type: none"> <li>● The PGP target goals are loosely connected to quantitative and/or qualitative data; changes in student performance cannot be explained by changes in professional practice.</li> </ul>	<p>This cell left intentionally blank</p>

APPENDIX M  
PROFESSIONAL GROWTH PLAN DEVELOPMENT SCORING AND  
FEEDBACK FORM



### Professional Growth Plan (PGP) Development Scoring and Feedback

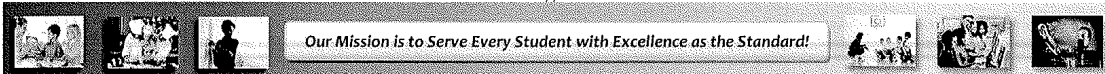
Teacher Number \_\_\_\_\_ School \_\_\_\_\_

Development Goal					
Professional Growth Plan: Development					
Elements (Refer to PGP Development Rubric)	Ratings & Scoring				Comments
	D(4)	P(3)	PSN(2)	U(0)	
1) Development of PGP Goal					
2) Work Plan Strategies					
Elements (Refer to PGP Development Rubric)	Ratings & Scoring				Comments
	D(2)	P(1)	PSN(0)	U(0)	
3) Outcome Measures & Reflections					
<i>column subtotals</i>					

Total score:

\_\_\_\_\_/\_\_\_\_\_  
Evaluator's Number      Date

IPPAS Appendix: 32



APPENDIX N  
PROFESSIONAL GROWTH PLAN DEVELOPMENT  
ADMINISTRATOR FINAL SCORING AND FEEDBACK FORM



Professional Growth Plan (PGP) Development  
Administrator Final Scoring and Feedback

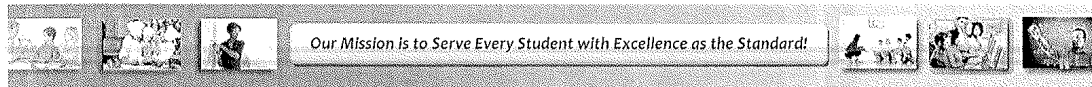
Teacher Number \_\_\_\_\_ School \_\_\_\_\_

Development Goal					
Professional Growth Plan: Development					
Elements (Refer to PGP Development Rubric)	Ratings & Scoring				Comments
	D(4)	P(3)	PSN(2)	U(0)	
1) Development of PGP Goal					
2) Work Plan Strategies					
Elements (Refer to PGP Development Rubric)	Ratings & Scoring				Comments
	D(2)	P(1)	PSN(0)	U(0)	
3) Outcome Measures & Reflections					
<i>column subtotals</i>					

Total Administrator Score:

Professional Growth Plan Development Final Scoring						
Professional Growth Plan (PGP) Development	Teacher Leader Scores				Admin Score	Averaged Pts Earned
	(1)	(2)	(3)	Average		

\_\_\_\_\_/\_\_\_\_\_  
Administrator Date





APPENDIX O  
MID-YEAR CONFERENCE FORM



Mid-Year Conference

Teacher Name \_\_\_\_\_ School \_\_\_\_\_

General Reflection

How are you progressing on your PGP goals?

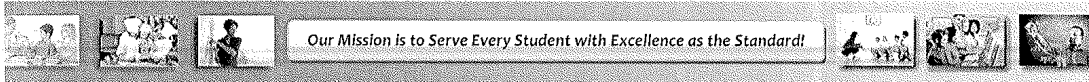
Empty box for reflection on PGP goals.

Are you meeting or not meeting the goals you established?

Empty box for reflection on goal meeting.

\_\_\_\_\_/\_\_\_\_\_  
Employee's Signature Date  
(Blue Ink Only)

\_\_\_\_\_/\_\_\_\_\_  
Administrator's Signature Date  
(Blue Ink Only)



APPENDIX P  
INDIVIDUAL PROFESSIONAL GROWTH PLAN IMPLEMENTATION  
RUBRICS

## Individual Professional Growth Plan (PGP) Implementation

**Plan Implementation** – this section is to be completed by the school principal (or designee) and the teacher who designed the Professional Growth Plan. The administration will monitor the implementation of the PGP, provide feedback and use rubric below for determining the level of implementation. Teacher will also self-assess and score their level of implementation as indicated below. The administration and teacher will meet to discuss progress and/or determine final scoring.

Elements	Distinguished (4 pts)	Proficient (3 pts)	Professional Support Needed (2 pts)	Unsatisfactory (0 pts)
<b>Working the Plan:</b> Implements the PGP with fidelity and professional practice	<ul style="list-style-type: none"> <li>● Fidelity is consistently evident both in the teacher's participation in the identified professional development and in the implementation of professional practice strategies.</li> <li>● New practice is sustained throughout lessons and is consistently linked to earlier, formative attempts or feedback.</li> <li>● Timeline(s) is/are followed and/or adjusted with appropriate rationale.</li> </ul>	<ul style="list-style-type: none"> <li>● Teacher consistently participates in professional development.</li> <li>● While new instructional strategies are implemented, they are inconsistently integrated into lessons.</li> <li>● Timelines are followed and/or adjusted with appropriate rationale.</li> </ul>	<ul style="list-style-type: none"> <li>● Teacher is marginally engaged in professional development.</li> <li>● New instructional strategies are implemented inconsistently.</li> <li>● Timelines are inconsistently followed.</li> <li>● Later practice is inconsistently linked to earlier attempts.</li> </ul>	<ul style="list-style-type: none"> <li>● Little or no evidence teacher is engaged in professional development and/or implementing newly acquired instructional strategies.</li> <li>● Timelines are either not established or not followed.</li> <li>● There are disconnects between earlier attempts and later practice.</li> </ul>
<b>Working the Plan:</b> Seeks feedback and support and shares successful practice	<ul style="list-style-type: none"> <li>● Feedback and support are actively sought and implemented.</li> <li>● Successful practice is readily shared and teacher actively seeks opportunities to do so.</li> <li>● Teacher has evidence of at least three collegial observations for the PGP Implementation.</li> <li>● Evidence that instruction has been modified based on feedback from the colleagues.</li> </ul>	<ul style="list-style-type: none"> <li>● Seeks feedback and support.</li> <li>● Will share successful practice when asked to do so.</li> <li>● Teacher has evidence that one to two colleagues have observed the implementation of identified strategies and that the teacher has modified instruction based on the feedback from colleagues.</li> </ul>	This cell left intentionally blank	<ul style="list-style-type: none"> <li>● Feedback has not been sought.</li> <li>● There is no evidence feedback was utilized to modify instruction.</li> <li>● Successful practice is not shared.</li> </ul>
<b>In-Process Monitoring</b>	<ul style="list-style-type: none"> <li>● Reflection and appropriate modifications to instructional techniques for continuous improvement are clearly evident.</li> <li>● Defined in-process measurement elements were consistently utilized.</li> <li>● In-process efforts provided varied and ongoing formative data related to student mastery and success.</li> </ul>	<ul style="list-style-type: none"> <li>● Reflection is sporadic.</li> <li>● Defined in-process measurement elements not consistently utilized.</li> <li>● In-process efforts provided ongoing formative data related to student mastery and success.</li> </ul>	This cell left intentionally blank	<ul style="list-style-type: none"> <li>● No evidence of reflection.</li> <li>● No defined in-process measurement elements.</li> <li>● Little or no evidence of on-going formative data related to student mastery and success.</li> </ul>

APPENDIX Q  
PROFESSIONAL GROWTH PLAN IMPLEMENTATION AND  
FEEDBACK FORM



Professional Growth Plan (PGP) Implementation
Scoring and Feedback

Teacher

School

Administrator

Development Goal

Professional Growth Plan Implementation: Working the Plan

Table with 4 columns: Elements (Refer to PGP Implementation Rubric), Self Ratings & Scoring, Teacher Comments, and Administrator Ratings & Scoring. Contains two rows of goal elements.

Professional Growth Plan Implementation: In-Process Monitoring

Table with 4 columns: Elements (Refer to PGP Implementation Rubric), Self Ratings & Scoring, Teacher Comments, and Administrator Ratings & Scoring. Contains one row for in-process monitoring.

column subtotals

column subtotals

Teacher Score: [ ]

Administrator Score: [ ]

Total Averaged Score: [ ]

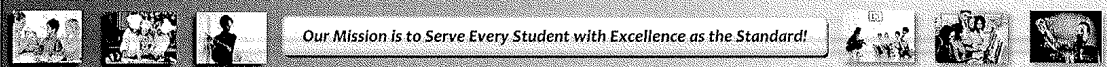
Administrator Comments:

Four horizontal lines for entering administrator comments.

Employee's Signature / Date

Administrator's Signature / Date

IPPAS Appendix: 34



APPENDIX R  
COLLABORATION & MUTUAL ACCOUNTABILITY TEAM  
SCORING FORM



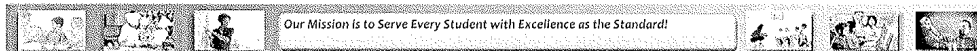
Collaboration & Mutual Accountability Team Scoring Form

School/Dept Name: Sunrise Elementary S/D #: 2221
Collaboration Team Member: Doe, Jane A. Date: 1/31/2012
Collaboration Team #: 2221-0001 (School number + 4 numeric digits defined by school. i.e.: 20220001)

Table with 5 columns: 8 Points, 6 Points, 4 Points, 2 Points, 0 Points. Rows describe teacher behaviors such as seeking input from colleagues, interacting with colleagues, and participating in school initiatives.

Table with 3 columns: Team Member Name, Points, Comments (Optional). Header row contains 'Last, First M'. Below are multiple empty rows for data entry.

Collaboration Team Member Signature \_\_\_\_\_





APPENDIX S  
VALUE-ADDED MEASURES



### Value-Added Measures

Elementary VAM Cut Scores

# Points	VAM Score Range	% included (prev 3 yrs)
35	0.96-2.50	2.4%
34	0.81-0.95	1.5%
33	0.66-0.80	2.8%
32	0.51-0.65	4.2%
31	0.36-0.50	5.2%
30	0.21-0.35	13.3%
29	0.06-0.20	27.5%
28	-0.09-0.05	27.6%
27	-0.24-0.10	10.9%
26	-0.39-0.25	3.2%
25	-0.54-0.40	0.8%
24	-0.69-0.55	0.4%
23	-0.84-0.70	0.0%
22	-0.99-0.85	0.1%
21	-1.14-1.00	0.1%
20	-1.29-1.15	0.00%
19	-1.44-1.30	0.0%

Middle VAM Cut Scores

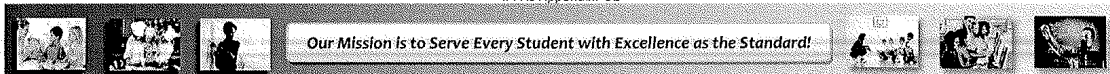
# Points	VAM Score Range	% included (prev 3 yrs)
35	0.91-1.10	0.0%
34	0.76-0.90	0.0%
33	0.61-0.75	0.3%
32	0.46-0.60	0.0%
31	0.31-0.45	0.3%
30	0.16-0.30	2.4%
29	0.01-0.15	25.7%
28	-0.14-0.00	50.3%
27	-0.29-0.15	17.8%
26	-0.44-0.30	2.4%
25	-0.59-0.45	0.7%
24	-0.74-0.60	0.0%

High School VAM Cut Scores

# Points	VAM Score Range	% included (prev 3 yrs)
35	0.91-3.50	3.3%
34	0.76-0.90	0.2%
33	0.61-0.75	1.3%
32	0.46-0.60	1.3%
31	0.31-0.45	3.5%
30	0.16-0.30	7.4%
29	0.01-0.15	35.4%
28	-0.14-0.00	29.3%
27	-0.29-0.15	4.6%
26	-0.44-0.30	3.5%
25	-0.59-0.45	2.6%
24	-0.74-0.60	3.5%
23	-0.89-0.75	1.3%
22	-1.04-0.90	1.3%
21	-1.19-1.05	0.2%
20	-1.34-1.20	0.0%
19	-1.49-1.35	0.4%
18	-1.64-1.50	0.0%
17	-1.79-1.65	0.2%
16	-1.94-1.80	0.0%
15	-2.09-1.95	0.2%
14	-2.24-2.10	0.4%
13	-2.39-2.25	0.0%
12	-2.54-2.40	0.0%
11	-2.69-2.55	0.2%

Alt School VAM Cut Scores

# Points	VAM Score Range	% included (prev 3 yrs)
35	0.91-1.10	0.0%
34	0.76-0.90	0.0%
33	0.61-0.75	0.0%
32	0.46-0.60	0.0%
31	0.31-0.45	0.0%
30	0.16-0.30	0.0%
29	0.01-0.15	0.0%
28	-0.14-0.00	83.3%
27	-0.29-0.15	16.7%



## REFERENCES

- American Institute for Research. (2011, August 1-2). Florida's value-added technical assistance workshop [PowerPoint]. Retrieved from <http://www.fldoe.org/Committees/pdf/august12tampres.pdf>
- Aunchman, D. S. (2009). *An investigation of the relationship between an administrator's perceived effectiveness of a teacher and value-added assessment in the elementary schools in the state of Florida* (Doctoral dissertation, Oral Roberts University). Retrieved from <http://search.proquest.com.ezproxy.lib.ucf.edu/docview/305185359?accountid>
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the use of student scores to evaluate teachers* (Working Paper No. 278). Retrieved from the Economic Policy Institute website: <http://www.epi.org/page/-/pdf/bp278.pdf>
- Barnatt, J. (2008, March). *But are they learning? Teachers candidates' use of inquiry focused on student outcomes*. Paper presented at AERA, New York. Retrieved from [http://tne.bc.edu/documents/Paper\\_1\\_Barnatt.pdf](http://tne.bc.edu/documents/Paper_1_Barnatt.pdf)
- Barrette, C., Morton, E., & Tozcu, A. (1995). *An overview of teacher evaluation* (Working paper). Retrieved from the University of Arizona website: <http://www.coh.arizona.edu/AWP/AWP3/AWP3%5Bbarrette%5D.pdf>
- Baum, D. (2011). John Tyler: Learning about teacher evaluation systems. Retrieved from Brown University website: <http://news.brown.edu/features/2011/04/tyler>
- Berry, B., Daughtrey, A., & Wieder, A. (2009, December). *Collaboration: Closing the effective teaching gap* (Policy Brief). Retrieved from the Center for Teaching Quality website: [http://teachersnetwork.org/effectiveteachers/images/CTQPolicyBriefOn\\_COLLABORATION\\_\\_021810.pdf](http://teachersnetwork.org/effectiveteachers/images/CTQPolicyBriefOn_COLLABORATION__021810.pdf)
- Black, S. (1993). How teachers are reshaping evaluation procedures. *Educational Leadership*, 51(2), 38-42.
- Bolam, R., McMahon, A., Stoll, L., Thomas, S., & Wallace, M. (2005). *Creating and sustaining professional learning communities* (Research report No. 637). London, England: General Teaching Council for England, Department for Education and Skills. Retrieved from [http://groups.ascd.org/resource/documents/133180-Bolam\\_Stoll\\_Creating\\_Sustaining\\_PLC.pdf](http://groups.ascd.org/resource/documents/133180-Bolam_Stoll_Creating_Sustaining_PLC.pdf)
- Borman, G. D., & Kimball, S. M. (2004). *Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student*

*achievement gaps?* (Working Paper No. TC-04-03). Retrieved from Consortium for Policy Research in Education website: [http://cpre.wceruw.org/papers/teacher\\_equity\\_aera04.pdf](http://cpre.wceruw.org/papers/teacher_equity_aera04.pdf)

- Borthwick, E., Cohodes, S., Sennette, J., & Touhey, A. (2009). *History of teacher evaluation models in American public education* (Report 4060867). Retrieved from Harvard Graduate School of Education website: <http://a100educationalpolicy.pbworks.com/w/page/3764950/Teacher%20Evaluation%20Models>
- Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service. Retrieved from <http://www.isbe.state.il.us/peac/pdf/primer.pdf>
- Brevard Public Schools. (2011). *Instructional personnel performance appraisal system*. Retrieved from <http://benefits.brevard.k12.fl.us/HR/comp/pas/pinfo/IPPAS.pdf>
- Brownell, M. T., Yeager, E., Rennells, M. S., & Riley, T. (1997). Teachers working together: What teacher educators and researchers should know. *Teacher Education and Special Education*, 20(4), 340-359. Retrieved from EBSCOhost.
- Burke, K. (2000). Results-based professional development. *NASSP Bulletin*, 84(618), 29-37.
- Cameron, A. (2011). *Professional growth plans for clock hours: Effects of Washington state education policy on student outcomes* (Master's thesis). Retrieved from [https://digital.lib.washington.edu/dspace/bitstream/handle/1773/16631/A.Cameron\\_Final\\_Capstone\\_Manuscript.pdf?sequence=1](https://digital.lib.washington.edu/dspace/bitstream/handle/1773/16631/A.Cameron_Final_Capstone_Manuscript.pdf?sequence=1)
- City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education: A network approach to improving teaching and learning*. Cambridge, MA: Harvard Education Press.
- Cochran-Smith, M., Barnatt, J., Friedman, A., & Pine, G. (2009). Inquiry on inquiry: Practitioner research and student learning. *Action in Teacher Education*, 31(2), 17-32.
- Cochran-Smith, M., & Lytle, S. L. (1999). The teacher research movement: A decade later. *Educational Researcher*, 28(7), 15-25.
- Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen-Vogel, L. (2005). Federal role in teacher quality: "Redefinition" or policy alignment? *Educational Policy*, 19(1), 18-43. doi: 10.1177/0895904804272246

- Connor, W. L. (1920). A new method of rating teachers. *The Journal of Educational Research*, 1(5), 338-358. Retrieved from <http://www.jstor.org/stable/27524324>
- Corcoran, S. P., Jennings, J. L., & Beveridge, A. A. (2011). *Teacher effectiveness on high- and low-stakes tests*. Retrieved from [https://files.nyu.edu/sc129/public/papers/corcoran\\_jennings\\_beveridge\\_2011\\_wkg\\_teacher\\_effects.pdf](https://files.nyu.edu/sc129/public/papers/corcoran_jennings_beveridge_2011_wkg_teacher_effects.pdf)
- Croft, A., Coggshall, J. G., Dolan, M., Powers, E., & Killion, J. (2010) *Job-embedded professional development: What it is, who is responsible, and how to get it done well* (Issue Brief). Retrieved from National Comprehensive Center for Teacher Quality website: [http://www.tqsource.org/publications/JEPD%20Issue %20 Brief.pdf](http://www.tqsource.org/publications/JEPD%20Issue%20Brief.pdf)
- Daley, G., & Kim, L. (2010). *A teacher evaluation system that works* (Working paper). Report from National Institute for Excellence in Teaching. Retrieved from [http://www.tapsystem.org/publications/wp\\_eval.pdf](http://www.tapsystem.org/publications/wp_eval.pdf)
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional learning*. Princeton, NJ: Educational Testing Service.
- Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappan*, 78(3), 193-200.
- Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world*. San Francisco, CA: John Wiley & Sons, Inc. Retrieved from [http://books.google.com/books?id=H0uUGKrESDUC&pg=PA486&source=gbs\\_selected\\_pages&cad=3#v=onepage&q&f=false](http://books.google.com/books?id=H0uUGKrESDUC&pg=PA486&source=gbs_selected_pages&cad=3#v=onepage&q&f=false)
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality* (Executive summary). Retrieved from the Center for American Progress website: [http://www.americanprogress.org/issues/2009/06/pdf/teacher\\_evaluation.pdf](http://www.americanprogress.org/issues/2009/06/pdf/teacher_evaluation.pdf)
- DuFour, R. (2004). What is a professional learning community? *Educational leadership*, 61(8), 6-11. Retrieved from [http://pdonline.ascd.org/pd\\_online/secondary \\_reading/ el200405\\_dufour.html](http://pdonline.ascd.org/pd_online/secondary_reading/el200405_dufour.html)
- Dunaway, D. M., Do-Hong, K., & Szad, E. R. (2010). Perceptions of the purpose and value of required individual growth plans. *The Educational Forum*, 74(1), 4-18.

- Englert, C. S., Tarrant, K. L., & Rozendal, M. S. (1993). Educational innovations: Achieving curricular change through collaboration. *Education and Treatment of Children*, 16(4), 441-473. Retrieved from EBSCOhost.
- Florida Department of Education. (2011). *What is the FCAT?* Retrieved from <http://www.fldoe.org>
- Florida Department of Education. (Producer). (2011). *Florida's value added model* [PowerPoint]. Available from [http://www.fldoe.org/committees/pdf/PresentationValue-added Model.pdf](http://www.fldoe.org/committees/pdf/PresentationValue-addedModel.pdf)
- Florida Department of Education (2012). *Recommendations of the Florida student growth implementation committee: Background and summary*. Retrieved from <http://www.fldoe.org>
- Florida Department of Education (2012). *2012 algebra I and geometry EOC assessment fact sheets*. Retrieved from <http://fc.fldoe.org/eoc/>
- Florida Government. (2011). *Governor Scott signs student success act*. Retrieved from <http://www.flgov.com/2011/03/24/governor-scott-signs-student-success-act/>
- Fullan, M. G. (1985). Change processes and strategies at the local level. *Elementary School Journal*, 85, 391-421. Retrieved from EBSCOhost.
- Fullan, M. G. (2001). *The new meaning of educational change*. New York: Teachers College Press.
- Fulton, K., & Britton, T. (2011). *STEM teachers in professional learning communities: From good teachers to great teaching*. Retrieved from the National Commission on Teaching and America's Future website: <http://nctaf.org/wp-content/uploads/2012/01/NCTA>
- Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79-107.
- Geringer, J. (2003). Reflections on professional development: Toward high-quality teaching and learning. *Phi Delta Kappan*, 84(5), 373-375.
- Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D., Whitehurst, G. J., & Croft, M. (2011). *Passing muster: Evaluating teacher evaluation systems*. Retrieved from Brookings website: [http://www.brookings.edu/reports/2011/0426\\_evaluating\\_teachers.aspx](http://www.brookings.edu/reports/2011/0426_evaluating_teachers.aspx)
- Goddard, Y., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and

- student achievement in public elementary schools. *Teachers College Record*, 109(4), 877-896.
- Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. Report of the National Comprehensive Center for Teacher Quality. Retrieved from <http://www.tqsource.org/publications/LinkBetweenTQandStudentOutcomes.pdf>
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis* (Working paper). Retrieved from the National Comprehensive Center for Teacher Quality website: <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>
- Goldrick, L. (2002). *Improving teacher evaluation to improve teaching quality* (Issue Brief). Retrieved from the National Governor's Association for Best Practices website: <http://www.nga.org/files/live/sites/NGA/files/pdf/1202IMPROVINGTEACHEVAL.pdf>
- Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education*, 74(3), 251-266.
- Gordon, R. (Producer). (2008, May 23). *Value-added measures: Implications for policy & practice* [PowerPoint]. Retrieved from <http://www.caldercenter.org/events/upload/Gordon-VAM.ppt>
- Gruenert, S. (2005). Correlations of collaborative school cultures with student achievement. *NASSP Bulletin*, 89(645), 43-55. Retrieved from EBSCOhost.
- Guskey, T. R. (1997). Research needs to link professional development and student learning. *Journal of Staff Development*, 18, 36-40. Retrieved from EBSCOhost.
- Haithcock, F. (2011, January 7). [Memorandum to District School Superintendents]. Florida Department of Education, Tallahassee, FL.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *The Journal of Political Economy*, 100(1), 84-117. Retrieved from <http://edpro.stanford.edu>
- Hanushek, E. A. (2002). Teacher quality. In L. T. Izumi, & W. M. Evers (Eds.), *Teacher quality* (pp. 1-12). Palo Alto, CA: Hoover Press. Retrieved from <http://edpro.stanford.edu/hanushek/admin/pages/files/uploads/Teacher%20quality.Evers-Izumi.pdf>
- Harris, D. N. (2012). *How do value-added indicators compare to other measures of teacher effectiveness?* Retrieved from the Carnegie Knowledge Network website: <http://www.carnegieknowledgenetwork.org/briefs/value-added/value-added-other-measures/>

- Haycock, K. (1998). Good teaching matters a lot. *OAH Magazine of History*, 13(1), 61-63. Retrieved from <http://www.jstor.org.ezproxy.lib.ucf.edu>
- Haycock, K., & Hanushek, E. (2010). An effective teacher in every classroom. *Education Next*, 10(3), 46-52. Retrieved from <http://web.ebscohost.com.ezproxy.lib.ucf.edu/>
- Hazi, H. M., & Rucinski, D. A. (2009). Teacher evaluation as policy target: Viable reform venue or just another tap dance? *ERS Spectrum*, 27(3), 31-40. Retrieved from [http://portal.ers.org/search/item\\_preview.cfm?assetid=30960](http://portal.ers.org/search/item_preview.cfm?assetid=30960)
- Heneman, H. G., Milanowski, A., Kimball, S. M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (Policy brief). Retrieved from the University of Pennsylvania Consortium for Policy Research in Education's website: [http://www.cpre.org/images/stories/cpre\\_pdfs/RB45.pdf](http://www.cpre.org/images/stories/cpre_pdfs/RB45.pdf)
- Hill, C. W. (1921). The efficiency of teachers. *The Elementary School Journal*, 21(6), 438-443. Retrieved from <http://www.jstor.org.ezproxy.lib.ucf.edu/stable/pdfplus/993672.pdf?acceptTC=true>
- Hill, H. (2009). Fixing teacher professional development. *Phi Delta Kappan*, 90(7), 470-476.
- Hill, H. C., Kapitula, L., & Umland, K. (2010). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.
- Hinchey, P. H. (2010). *Getting teacher assessment right: What policymakers can learn from research*. Report of the National Education Policy Center. Retrieved from [http://greatlakescenter.org/docs/Policy\\_Briefs/Hinchey\\_TeacherEval.pdf](http://greatlakescenter.org/docs/Policy_Briefs/Hinchey_TeacherEval.pdf)
- Hollins, E. R., McIntyre, L. R., DeBose, C., Hollins, K. S., & Towner, A. (2004). Promoting a self-sustaining learning community: Investigating an internal model for teacher development. *International Journal of Qualitative Studies in Education*, 17(2), 247-264. Retrieved from EBSCOhost.
- Hughes, V. M. (2006). *Teacher evaluation practices and teacher job satisfaction* (Doctoral dissertation, University of Missouri). Retrieved from <https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/4468/research.pdf?sequence=3>
- Ishii, J., & Rivikin, S. G. (2009). Impediments to the estimation of teacher value added. *Finance and Policy*, 4(4), 520-537. Retrieved from EBSCOhost.
- Jackson, C., & Bruegmann, E. (2009, February). *Teaching students and teaching each other: The importance of peer learning for teachers* (Working Paper No. 77). Retrieved from the Cornell University, ILR School's website: <http://digital>



commons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1079&context=workingpapers&sei-redir=1#search=%22Jackson%2C%20Bruegmann%20Teaching%20students%20teaching%20each%20other%3A%20importance%20peer%20learning%20teachers%22

- Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
- Jerald, C. (2012). *Ensuring accurate feedback from observations: Perspectives on practice*. Retrieved from the Bill & Melinda Gates Foundation website: <http://www.gatesfoundation.org/college-ready-education/Documents/ensuring-accuracy-up.pdf>
- Kane, T. J. & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (Policy and Practice Brief). Retrieved from Bill & Melinda Gates Foundation website: [http://www.metproject.org/downloads/MET\\_Gathering\\_Feedback\\_Research\\_Paper.pdf](http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf)
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). *Identifying effective classroom practices using student achievement data* (Working paper 15803). Retrieved from the National Bureau of Economic Research website: [http://www.gse.harvard.edu/ncte/resources/publications/Identifying\\_Effective\\_Classroom\\_Practices\\_-\\_Kane\\_Taylor\\_Tyler\\_Wooten.pdf](http://www.gse.harvard.edu/ncte/resources/publications/Identifying_Effective_Classroom_Practices_-_Kane_Taylor_Tyler_Wooten.pdf)
- Katzenbach, J. R., & Smith, D. K. (1993). The discipline of teams. *Harvard Business Review*, 71(2), 111-120. Retrieved from EBSCOhost.
- Koedel, C., & Betts, J. R. (2007). *Re-examining the role of teacher quality in the educational production function* (Working paper 0708). Retrieved from [http://economics.missouri.edu/working-papers/2007/wp0708\\_koedel.pdf](http://economics.missouri.edu/working-papers/2007/wp0708_koedel.pdf)
- Koretz, D. (2008). A measured approach. *American Educator*, Fall, 18-39. Retrieved from <http://www.aft.org/pdfs/americaneducator/fall2008/koretz.pdf>
- Kovats, S. M. (2006). *The impact of the teacher evaluation process on teacher practice and student achievement in an urban school: A case study* (Doctoral dissertation, University of Southern California). Retrieved from <http://search.proquest.com.ezproxy.lib.ucf.edu/docvie/304967618?accountid=10003>
- Lawrence, G., & Florida State Dept. of Education, T. n. (1974). Patterns of effective inservice education. A state of the art Summary of research on materials and procedures for changing teacher behaviors in inservice education. Retrieved from EBSCOhost.

- Liu, C. W. (2010). The implementation of teacher evaluation for professional development in primary education in Taiwan (Doctoral dissertation). Retrieved from <http://www.bookpump.com/dps/pdf-b/9423804b.pdf>
- Lockwood, J. R., Louis, T. A., & McCaffrey, D. F. (2002). Uncertainty in rank estimation: Implications for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics*, 27(3), 255-270. Retrieved from EBSCOhost.
- Lomax, R.G. (2007). *An introduction to statistical concepts* (2<sup>nd</sup> ed.). New Jersey: Lawrence Erlbaum Associates.
- Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, 33, 757-798. Retrieved from EBSCOhost.
- Marzano, R., Toth, M., & Schooling, P. (2012). *Examining the role of teacher evaluation in student achievement: Contemporary research base for the Marzano causal teacher* (Working paper). Retrieved from the Learning Science Marzano Center website: <http://ok.gov/sde/sites/ok.gov.sde/files/TLE-MarzanoWhitePaper.pdf>
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67-101. Retrieved from EBSCOhost.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education, Finance and Policy*, 4(4), 572-606.
- McClellan, C. (2012). *Teacher evaluator training & certification: Lessons learned from the measures of effective teaching project*. Retrieved from [www.teachscape.com](http://www.teachscape.com)
- McLaughlin, M. W. (1990). The rand change agent study revisited: Macro perspectives and micro realities. *Educational Researcher*, 19(1), 11-16. Retrieved from EBSCOhost.
- McLaughlin, M. W., & Marsh, D. D. (1978). Staff development and school change. *Teachers College Record*, 47(8), 57-61. Retrieved from EBSCOhost.
- Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research*, 80(4), 242-247. Retrieved from <http://web.ebscohost.com.ezproxy.lib.ucf.edu/ehost/pdfviewer/pdfviewer?vid=4&hid=14&sid=7399976a-fa62-4c27-b624-c29554f785b1%40sessionmgr4>

- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33-53.
- Montalvo, G. P., Mansfield, E. A., & Miller, R. B. (2007). Liking or disliking the teacher: Student motivation, engagement and achievement, *Evaluation and Research in Education*, 20(3), 144-158.
- Munoz, M. A., & Chang, F. C. (2007). The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education*, 20(3-4), 147-164.
- Musanti, S. I., & Pence, L. (2010). Collaboration and teacher development: Unpacking resistance, constructing knowledge, and navigating identities, *Teacher Education Quarterly*, 37(1), 73-85.
- National Board for Professional Teaching Standards. (Producer). (2011, March 10). *Student learning, student achievement: How do teachers measure up?* [Webcast]. Retrieved from <http://dl.nmmstream.net/media/nbpts/flash/100311webcast/100311loader.html>
- The National Commission on Excellence in Education. (1983). *A nation at risk: The imperative of educational reform*. Retrieved from <http://teachertenure.procon.org/sourcefiles/a-nation-at-risk-tenure-april-1983.pdf>
- The New Teacher Project. (2010). *Teacher evaluation 2.0*. Retrieved from <http://tntp.org/files/Teacher-Evaluation-Oct10F.pdf>
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23), 1-23.
- NGA Center for Best Practices. (2009, October). *State policies to improve teacher professional development* (Issue Brief). Washington DC: Grossman, T. & Hirsch, E. Retrieved from <http://www.eric.ed.gov/PDFS/ED507644.pdf>
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237-257. Retrieved from EBSCOhost.
- Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163-193.

- Pate, G. Y. (2010). *The Texas professional development and appraisal system: Links to student achievement* (Doctoral dissertation, Capella University). Retrieved from <http://search.proquest.com.ezproxy.lib.ucf.edu/docview/839826028?accountid=1003>
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin Press.
- Peterson, K. D. (2004). Research on school teacher evaluation, *NASSP Bulletin*, 88(639), 60-79.
- Peterson, K. D., Wahlquist, C., Bone, K., Thompson, J., & Chatterton, K. (2001). Using more data sources to evaluate teachers. *Educational Leadership*, 58(5), 40-43.
- Phillips, J. (2003). *Powerful learning: Creating learning communities in urban school reform*. *Journal of Curriculum and Supervision*, 18(3), 240-258. Retrieved from EBSCOhost.
- Rosenholtz, S. J., & Simpson, C. (1990). Workplace conditions and the rise and fall of teachers' commitment. *Sociology of Education*, 63(4), 241-257.
- Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education*, 17(1), 51-65. Retrieved from EBSCOhost.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (Research progress report). Retrieved from <http://www.mccsc.edu/~curriculum/cumulative%20and%20residual%20effects%20of%20teachers.pdf>
- Sanders, W. L. (2006, October). *Comparisons among various educational assessment value-added models*. Paper presented at the Power of Two-National Value-Added Conference, Columbus, OH. Retrieved from <http://www.sas.com/resources/asset/vaconferencepaper.pdf>
- Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation* (Research Report). Retrieved from the University of Chicago Consortium on Chicago School Research website: <http://ccsr.uchicago.edu/publications/Teacher%20Eval%20Report%20Final.pdf>
- Sass, T. R. (2008). *The stability of value-added measures of teacher quality and implications for teacher compensation policy* (Brief No. 4). Washington, D.C.: Calder. Retrieved from [http://www.urban.org/uploadedpdf/1001266\\_stabilityofvalue.pdf](http://www.urban.org/uploadedpdf/1001266_stabilityofvalue.pdf)
- Schacter, J., & Thum Y. M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review*, 23(4), 411-430. Retrieved from EBSCOhost.

- Schafer, K. (2012, November 4). An overview of Brevard's 2012 instructional personnel performance appraisal system [Powerpoint]. Retrieved from [http://benefits.brevard.k12.fl.us/HR/comp/pas/ipinfo/2012\\_Evaluation\\_System\\_Overview\\_10.31.12.pdf](http://benefits.brevard.k12.fl.us/HR/comp/pas/ipinfo/2012_Evaluation_System_Overview_10.31.12.pdf)
- Schmoker, M. (2004). At odds: Strategic planning – learning communities at the crossroads – toward the best schools we've ever had. *Phi Delta Kappan*, 86(1), 84-88. Retrieved from EBSCOhost.
- Shaffer, J. G. (1990). *A study of the relationship between student achievement and summative teacher evaluations* (Doctoral dissertation, East Texas State University). Retrieved from <http://search.proquest.com.ezproxy.lib.ucf.edu/dissertations/docview/303875315/fulltextPDF/130E79299F432B63BF2/1?accountid=10003>
- Shurtleff, D. S. & Lored, J. (2008). *Beyond no child left behind: Value-added assessment of student progress* (Brief Analysis No. 636). Retrieved from National Center for Policy Analysis website: <http://www.ncpa.org/pub/ba636>
- Smylie, M. A., Lazarus, V., & Brownlee-Conyers, J. (1996). Instructional outcomes of school-based participative decision making. *Educational Evaluation and Policy Analysis*, 18(3), 181-198.
- Sparks, D., & Loucks-Horsley, S. (1989). Five models of staff development. *Journal of Staff Development*, 10(4), 40-57.
- Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. *The Elementary School Journal*, 104(2), 127-146. Retrieved from EBSCOhost.
- Strong, M. (2011). *The highly qualified teacher*. New York: Teachers College Press.
- Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation*. Larchmont, NY: Eye on Education.
- Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record*, 104(8), 1591-1626. Retrieved from EBSCOhost.
- Teachers College Columbia University. (Producer). (2009, October 22). *Arne Duncan address* [Video]. Retrieved from <http://blip.tv/teachers-college-columbia-university/arne-duncan-speaks-at-teachers-college-oct-22-2009-2822240>
- Thornton, S. (2012). *The importance of reliability in performance appraisals*. Retrieved from [http://www.ehow.com/about\\_5445066\\_importance-reliability-performance-Appraisals.html](http://www.ehow.com/about_5445066_importance-reliability-performance-Appraisals.html)

- Tucker, D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tye, K.A., & Tye, B. B. (1984). Teacher isolation and school reform. *Phi Delta Kappan*, 65(5), 319-322. Retrieved from EBSCOhost.
- Tyler, J. H. (2011). *Designing high quality evaluation systems for high school teachers: Challenges and potential solutions*. Retrieved from [www.americanprogress.org](http://www.americanprogress.org)
- U.S. Department of Education. (2001). *Public law print of PL 107-110, the no child left behind act of 2001*. Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/index.html>
- U.S. Department of Education. (2009). *Race to the top program executive summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education. (2010). *A blueprint for reform: The reauthorization of the elementary and secondary education act*. Retrieved from <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf>
- Vevea, B. (2012). *Training vital to success of new teacher: Evaluation expert Charlotte Danielson spoke to educators, school leaders Tuesday*. Retrieved from <http://www.wbez.org/news/training-vital-success-new-teacher-evaluations-98338>
- Wallace, M. R. (2009). Making sense of the links: Professional development, teacher practices, and student achievement. *Teachers College Record*, 111(2), 573-596.
- Waxman, H. C. (2011). In StateUniversity.com Education Encyclopedia, *Classroom observation – purposes of classroom observation, limitations of classroom observation, new directions*. Retrieved from <http://education.stateuniversity.com/pages/1835/Classroom-Observation.html>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect. Our national failure to acknowledge and act on differences in teacher effectiveness* (Executive summary). Retrieved from The New Teacher Project website: [http://tntp.org/files/TheWidgetEffect\\_execsummary\\_2nd\\_ed.pdf](http://tntp.org/files/TheWidgetEffect_execsummary_2nd_ed.pdf)