

2013

## Classroom observation, self-assessment of efficacy, and student perceptions of engagement as predictors of value-added scores

Leslie Davis Blanchard

*Louisiana State University and Agricultural and Mechanical College*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_dissertations](https://digitalcommons.lsu.edu/gradschool_dissertations)



Part of the [Human Resources Management Commons](#)

---

### Recommended Citation

Blanchard, Leslie Davis, "Classroom observation, self-assessment of efficacy, and student perceptions of engagement as predictors of value-added scores" (2013). *LSU Doctoral Dissertations*. 1575.

[https://digitalcommons.lsu.edu/gradschool\\_dissertations/1575](https://digitalcommons.lsu.edu/gradschool_dissertations/1575)

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

CLASSROOM OBSERVATION, SELF-ASSESSMENT OF EFFICACY, AND  
STUDENT PERCEPTIONS OF ENGAGEMENT AS PREDICTORS OF VALUE-  
ADDED SCORES

A Dissertation

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

in

The School of Human Resource Education  
and Workforce Development

by

Leslie Davis Blanchard  
B.A. Nicholls State University, 1996  
M.Ed. Southern University Baton Rouge, 2001  
December 2013

This work, along with the whole of my life, is dedicated to my daughters. Thank you for being so flexible for the last several years. Thank you for understanding that I couldn't always be there at every one of your events. Thank you for always supporting me and for always believing in me. Please know that everything that I've ever done, every challenge I've ever tackled, and every decision I've ever made has had your ultimate well-being at its root. Cassidy Anne and Caroline Elizabeth Blanchard, I love you with my entire heart. Mommy's home.

## **ACKNOWLEDGEMENTS**

There are a great many individuals that deserve acknowledgement for their contribution to this accomplishment. In fact, there is simply no way in the world that this endeavor could have been undertaken without them. My professors supported me as a student while always understanding that I was also a mother and an employee; my family supported me as a mother while always understanding that I was also a student and an employee; my bosses and colleagues supported me as an employee while always understanding that I was also a mother and a student. Without this triad of mutual encouragement, my failure was imminent.

First and foremost, abundant thanks go out to the excellent academic support that I've received from my Louisiana State University committee. To my co-chair, Dr. Michael Burnett, and to Dr. Tracey Rizzuto and Dr. Pamela Blanchard, thank you for your excellent guidance, professional integrity, and specialized wisdom. Very special additional thanks to Dr. Eugene Kennedy for your assistance in the data and statistics functions. Your feedback and critical analysis have been invaluable, and I am very fortunate to have had such an esteemed committee to steer my research.

Sincere appreciation is also extended to the faculty, staff, and students of the "best school in the state", known here as Knightsville Academy. I've received wholehearted support from this district all the way from the executive levels of leadership down to the students in every class, and I could never have found a better case study school if I looked far and wide.

Next, I'd like to thank my family. My mom and dad have supported me in so many tangible and intangible ways. I can't begin to let them know how much I appreciate everything that they've contributed to my survival, and how it is my goal to always make them proud of me. My daughters have also been more than encouraging, and have demonstrated patience and understanding throughout the entire process when faced with "Mommy has homework", "Mommy has midterms", and "Mommy has class." Gregory Sevcik Jr. is my champion and my knight, and I can never let him know the impact of his belief in me. Thank you all from the bottom of my heart for making the machine that is my life function properly. I love you.

To my major professor, committee chair, intellectual mentor, and friend Chancellor William Richardson, I offer thanks at a level that mere words cannot convey. Dr. Richardson became my biggest academic fan when I walked into his first class in the fall of 2011, and he's been my most adamant scholastic cheerleader since then. He's modified his extraordinarily hectic schedule to include service as the chair of this committee, and has tolerated more than his fair share of accommodations to demonstrate support in levels that go far above and beyond that which a doctoral student should ever expect. He has provided professional opportunities and exposure that I couldn't have achieved on my own, and his transformational vision is the stuff of legends. Dr. Richardson's faith in my abilities and his confidence in my work have already propelled me into areas that I didn't think possible. The man is an icon in the eyes of many, and I

hope to forever be a small part of his educational legacy and life's mission to show my gratitude for his efforts on my behalf.

And finally, I humbly and honorably offer my most sincere thanks and my deepest appreciation to the one man without whom none of what I do would be doable. This man walked into my office four years ago, and he not only changed my career- he changed my life. What I knew at the time was that I was going to manage a Masters Degree program for teachers at LSU, and after a year I'd get to go to school there for free. What I could never have known was that I was going to be awed and inspired by a leader and a visionary whose idealistic beliefs in what teachers can do and what education can become is nothing short of noble. What I could never have known was that he would give me more support, more encouragement, more freedom, and more time than he would ever be expected to give. What I could never have known was that his generous nature, his kind demeanor, his brilliant mind, and his servant's heart would be so motivational. What I could never have known was that I wasn't only getting a new boss; I was getting the opportunity to work daily with an example of exactly the kind of leader I want to be. Dr. James J. Madden, thank you for everything that you've done for me. I'll never ever forget it.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	viii
ABSTRACT .....	ix
CHAPTER 1 INTRODUCTION .....	1
Rationale .....	1
Traditional Teaching Evaluation through Observation is a Flawed Process .....	3
Statement of the Problem .....	6
Purpose and Objectives of the Study .....	6
Significance of the Study .....	7
CHAPTER 2 REVIEW OF RELATED LITERATURE .....	10
History of American Education .....	10
The Purposes of Teacher Evaluation .....	13
History of Teacher Evaluation Practices .....	16
The Evaluation of Teaching is Essential to Effective Education .....	20
Traditional Teacher Evaluation .....	22
Current Status & Recent Developments in Teacher Evaluation in America .....	24
Current Status & Recent Developments in Teacher Evaluation Internationally ..	26
The Value Added Model .....	29
The Importance of Feedback, Trust and Follow-Up .....	32
360 Degree Feedback Models in other Professions .....	34
Student Feedback .....	36
Teacher Self-Efficacy .....	38
The MET Project .....	41
CHAPTER 3 METHODOLOGY .....	43
Research Design .....	43
Population and Sample .....	44
Instrumentation and Data Collection .....	45
Phase 1: Self-Efficacy Survey Administration .....	45
Phase 2: Demographic Data Collection .....	47
Phase 3: Student Survey Perceptions Administration .....	48
Phase 4: Principal Observation Scores Data Collection .....	50
Phase 5: Value Added Measures Scores Data Collection .....	51
Data Analysis .....	51
CHAPTER 4 PRESENTATION OF RESULTS .....	56
Overview .....	56
Summary of Response Rates and Total Sample .....	57
Results Addressing Research Hypotheses .....	60
Research Hypothesis 1 .....	60

Research Hypothesis 2 .....	61
Research Hypothesis 3 .....	62
Research Hypothesis 4 .....	63
Additional Results.....	63
CHAPTER 5 CONCLUSIONS.....	65
Summary .....	65
Implications, Recommendations, and Limitations .....	65
Conclusion 1.....	66
Conclusion 2.....	68
Conclusion 3.....	69
Additional Results.....	71
Limitations .....	71
BIBLIOGRAPHY .....	75
APPENDIX A CONSENT FORM .....	84
APPENDIX B PERMISSION TO USE TSES .....	85
APPENDIX C TEACHERS SENSE OF EFFICACY SCALE .....	86
APPENDIX D PERMISSION TO USE SIM INSTRUMENTS .....	87
APPENDIX E SCRIPTS .....	88
APPENDIX F STUDENT PERCEPTIONS INSTRUMENTS-SIM .....	89
APPENDIX G STUDENT PERCEPTIONS INSTRUMENTS-LSU .....	93
APPENDIX H TAP OBSERVATION RUBRIC.....	98
APPENDIX I COMPASS ACCOUNTABILITY MATRIX .....	107
APPENDIX J IRB APPROVAL.....	110
VITA .....	111

## **LIST OF TABLES**

Table 3.1 Reliabilities of the Teachers Sense of Efficacy Scale

Table 4.1 Demographics of Respondents used in Data Analysis

Table 4.2 Item by item analysis, grades 9-12 and SLT Scores

## **ABSTRACT**

The purpose of this study was to determine which has the strongest correlation to student achievement as measured by value-added test scores: Principal Observations, Teachers Self-Efficacy Ratings, or Student Perceptions of Teacher Effectiveness. 68 teachers from a K-12 public school in the southeast region of the United States agreed to participate in the study.

The Teachers' Sense of Efficacy Scale (TSES) was utilized to measure the teachers in terms of their own reports of self efficacy. The School Improvement Model (SIM) of Iowa State University instruments were used to measure the students' perceptions of teacher effectiveness. The Teacher Advancement Program (TAP) model observation protocol was used to measure the principals/supervisors' observation scores. These three measures were run in a multiple regression correlation to determine which of the three was the strongest predictor of student outcomes.

An analysis of Pearson's Moment Correlation among all three variables revealed that the principal observation scores were the only statistically significantly correlated measures that could be inferred to have any predictive impact on student achievement as measured by Value Added Scores.

## **CHAPTER 1: INTRODUCTION**

### **Rationale**

The quality of the teacher in a classroom is the most critical factor in the achievement of students. When taught by a “master teacher,” an individual student’s learning is extended an average of two to three additional months’ in math and reading compared with students taught by an average teacher, and five to six months further than compared to students taught by low-performing teachers. Teacher quality can also have lasting effects on students’ lives beyond academic performance. These types of gains during the formative years for children indicate a greater interest in attending college, higher employment and salaries as adults, and lower incidence of teenage pregnancy (TNTP, 2012). The importance, meaning, and irreplaceable nature of quality teaching performance are well documented by researchers (Donaldson, 2012; Isore’, 2009; McGuinn, 2012; Keeling, Mulhern, Sexton, & Weinberg, 2009; NCTQ Report, 2012; OECD, 2013; et al).

Despite the well-researched and supported importance of teacher quality, efforts to enhance teacher quality through different iterations of performance appraisals and evaluation have made very scant headway. The topic of teacher evaluation systems is one that sparks animated discussion at the mere mention of it, and few actors in the process are without opinions on the details. The issue of who should be included in the evaluation of a teacher is one of the major details. Others include which criteria should set the benchmarks for achievement, and how the results should be used.

The results of traditional evaluation programs based on principal observation have been generally ignored in terms of informing professional development, enhancing strategic improvement, improving the quality of instruction, or dismissing teachers who are not performing up to standards (Donaldson, 2009; Weisberg et al, 2009). The results are largely negligible at the other end of the quality spectrum, also, and rarely are exemplary teachers recognized or rewarded for their superior contributions (Sykes and Winchell, 2010).

Since the ultimate goal of teaching is to improve student learning, it is tempting to assess the quality of teaching performance in terms of student outcomes. Much research over several decades has focused on the use of student achievement to evaluate teachers' performance, and it is still a contentious issue. For most of these years, standardized test scores were used as an absolute in a vacuum, and a teachers' performance was measured based upon the performance of students at a given criterion achievement level (Isore', 2009). More recently, the evolution of various "value-added" models has gained popularity. Such models are specifically designed to statistically identify the contribution that an individual teacher has made---or the "value-added"--- to the students' cumulative learning. While this idea is initially attractive, it presents a great many caveats discussed further in Chapter 2. However, until a more desirable yardstick for outcome measure is created, test scores are the quintessential "dependent variable" for evaluation of teachers. Recent

developments in educational performance evaluation systems support this notion, including the Louisiana Compass Evaluation System.

### **Traditional Teaching Evaluation through Observation is a Flawed Process**

The faults that exist in most teacher evaluation models are many and glaring. The terms “professional growth” and “evaluation” are not interchangeable, and yet they are used that way in the vast majority of schools and other educational systems. Principals are considered to be educational leaders and in many cases do have the knowledge and the skills set required to impact instruction; however, they often have far too little time to be any more than building level managers. Observation visits are also scheduled far in advance, and teachers are aware of the hour and the class that will be evaluated. Glamorized lessons are often the result, and both teachers and principals agree that an atypical, planned performance is unrepresentative of what happens regularly in classrooms. Paperwork trails and write-ups are completed as mandates of the school, the district and/or the state, but these, too, are ceremonial in nature, and seldom have any weight or meaning beyond form-filing. They also do not focus on student learning, and instead rely largely on observable teacher behaviors that are expected to represent typical behaviors that always occur in the classroom. But with such a small amount of instructional time actually being observed, how can that generalization be made? The actual percentage of classroom teaching that administrators observe is minimal, and therefore an accurate analysis of actual daily classroom progress is impossible in the traditional framework. The belief that a one-shot 45-minute dog and pony

show can somehow be representative of daily teaching is foolishly optimistic and leads to inauthentic findings and inflated ratings of excellence. Dismissal for poor performance almost never occurs, in part because poor performance is almost never reported as a result of the observation.

Marshall has classified principals into three categories: the “saint” who follows the procedures of the observation process to the letter, often spending hours on each evaluation; the “cynic” who completes the minimum district requirements to comply with the evaluation mandate all the while maintaining the firm belief that they make little or no difference to teachers; and the “sinner” who rebelliously doesn’t do them at all (except occasionally in the event of an ineffective teacher) often citing time constraints and managerial responsibilities as excuses. When the true goal of the process is improving instruction, the saint is no more effective than the cynic or the sinner (Marshall, 2005).

These explanations should in no way be construed as an indictment of principals. The misconception that principals are too lazy or incompetent to evaluate correctly is fundamentally false and oversimplified. The deeper and more complex truth is that “administrators appear hesitant to assign negative ratings for many reasons – inadequate training, fear of pushback from faculty, and uncertainty about district support – that go well beyond a lack of will or follow-through” (Weisberg, 2009). Many new teacher evaluation frameworks include the assumption that principals will provide valuable feedback to their employees that is based on evidence of performance and student outcomes, and the principals’ ability to offer support and guidance for improvement is expected.

There is quite a disconnect between educational leadership preparation programs which train principals to be administrators and this expectation. Many principals feel ill-equipped and unprepared to give constructive and meaningful feedback, and unaware of what the content of that feedback should be (Goe, 2013).

Teachers find very little meaning or value in the traditional teacher evaluation process. Many factors contribute to this complete disregard. *The Widget Effect* reports that 39 percent of teachers in Denver public schools had a development area identified on their most recent evaluation, and yet “do not know” which performance standard they failed to meet (Weisberg et al, 2009). Standard, all-purpose instruments are devoid of meaning, and often give generic and vague feedback, if any. The credibility of the “drive-by” evaluation is also questionable because teachers are aware of the administrative pressures that detract from a principal’s ability to be thorough (Toch & Rothman, 2008).

Only a minute percentage of teachers claim that they have altered or changed their teaching practices in any way as a result of traditional principal evaluation. Kennedy (2005) investigated how teachers make improvements to their practice in *Inside Teaching: How Classroom Life Undermines Reform*. He saw that when teachers were unhappy with some aspect of their or their students’ performance or goal attainment, they tended to make small modifications to their routine over time. Huberman (1983) refers to this phenomenon as “experiential learning,” and it was found to have a much greater impact on teacher practice adjustment than evidence gained from “institutional

sources” like standardized tests, curriculum standards, district mandates, and textbooks or “knowledge vendors” like workshops, university courses, and membership in professional organizations (Kennedy, 2005). Experience drives adjustment; evaluation does not. However, when evaluation-for-accountability is once and for all separated from evaluation-for-improvement, the hope is that the use of meaningful feedback will influence instructional development.

### **Statement of the Problem**

The research problem of this study was “Which has the strongest correlation to student achievement as measured by value-added test scores: Principal Observations, Teachers Self-Efficacy Ratings, or Student Perceptions of Teacher Effectiveness?” To the knowledge of this author, no research studies have been conducted specifically comparing the correlation coefficients between these three measures in terms of student achievement for any individual case K-12 public school. This study sought to fill that void in the research.

### **Purpose and Objectives of the Study**

Both nationally and internationally, increased attention is being paid to the performance evaluation of teachers in efforts to improve teacher quality. Stakeholders, policy makers, administrators, and educators are confronted with elevated pressure to design and implement programs that evaluate teacher performance. Information regarding valuable and meaningful teacher evaluation plans is necessary to inform program decisions and influence policy. It is also desirable to finally have a policy that makes sense from the perspective of the

customer -the student- and gives voice to the eventual development of a true 360 degree feedback evaluation cycle for educators.

The purpose of this study was to compare three inputs of teacher evaluation to determine which is the most strongly correlated to the academic outcomes of students as measured by value-added test scores. The following list of specific objectives were developed to guide the researcher:

1. To determine if there is a statistically significant relationship between the principal's/supervisor's observation score of a teacher and the achievement of that teachers' students as measured by value-added standardized test scores.
2. To determine if there is a statistically significant relationship between the self-efficacy score of a teacher and the achievement of that teachers' students as measured by value-added standardized test scores.
3. To determine if there is a statistically significant relationship between the students' perceptions of engagement effectiveness of a teacher and the achievement of that teacher's students as measured by value-added standardized scores.

### **Significance of the Study**

The traditional measure of teacher evaluation is principal/supervisory observation, and was the sole means of performance measurement for many years. More recently, evaluation systems such as Louisiana's Compass Evaluation System have included a second variable of value-added test scores to

form a composite score for each teacher. A third variable of interest in the myriad of teacher evaluation is self-efficacy. In seminal research by psychologist Albert Bandura and others spanning more than 50 years, the interconnectedness and distinctions between what individuals feel that they CAN accomplish (efficacy beliefs) and judgments of the “likely consequences that behavior will produce” (outcome expectations) have been studied and argued (1986). People who genuinely expect to succeed in a specified endeavor tend to also expect successful outcomes from that endeavor. This statement is grossly oversimplified here, and explained in much greater detail in Chapter 2: Review of Related Literature. However for purposes of clarity, the study presented here investigated whether there is a significant relationship between the teachers’ sense of his/her own self-efficacy and the value-added student outcomes that are produced.

The fourth and final variable of interest in this study is the students’ perceptions of teacher effectiveness. This data collection stems from the perspective that the students in the classrooms are the “clients”, and that they are the audience most knowledgeable about the ongoing and regular performance of the teacher. As such, they should be able to provide an accurate assessment of the consistent performance of the teacher. Once the measures of student perception surveys were collected, they, too, were correlated to the value-added student outcomes from that teacher.

Data obtained from the completion of this study is intended to provide useful information to administrators, educators, policy-makers, parents, and students as to the relationships between rating inputs from various levels (top-

down, self-rate, and bottom-up) evaluation that can serve as a precursor to the eventual development of a true 360 degree evaluation program for teachers.

This study was proposed to contribute to the body of knowledge concerning the effectiveness of teacher evaluation programs. The findings can be beneficial to researchers and human resource professionals in school systems as they tailor their evaluation programs to meet the needs of their workforce. Informed professional development can be meaningfully derived from the data as teachers can identify strengths in their practice and areas where growth can occur.

## **CHAPTER 2: REVIEW OF RELATED LITERATURE**

### **History of American Education**

Education has been held as a primarily important function of society since the early years of the United States of America. In the first part of the Colonial period, the New England Colonies established school districts whose sole purpose was to expand educational opportunities to more people. Reading, writing and arithmetic were fundamental principles of learning and were often taught in the context of the Puritan Church. Interest in more practical application of education became popular in the middle colonies, which were characterized by many different kinds of religious denominations, and William Penn and Benjamin Franklin stressed such education toward the end of the colonial period. Townships for Education were established as a result of the Northwest Ordinance in 1787, which provided funds by the sale of land dedicated to the purpose of funding schools (Urban and Wagoner, 2004; Cremin, 1970).

In the 1840's, the common school movement—where education became available to the common masses instead of privatized endeavors—created an organized system of education. Reformers such as Horace Mann and Henry Barnard helped create statewide common school systems to increase opportunities for all children and to create common bonds among an increasingly diverse population. These reformers argued that expanding opportunities for education could enhance social stability and prevent crime and poverty. Advocates of the common school system worked to establish a free elementary education that would be accessible to everyone and financed by federal funds.

This movement resulted in the establishment of public schools that were accountable to local school boards and state governments. Compulsory school attendance laws for elementary aged children existed in all states by 1918. Public high schools were developed in the early 19<sup>th</sup> century as a similar federal institution alternative to the private academies of the 18<sup>th</sup> century (Tyack, 1967). In 1874, the Kalamazoo Case set a legal precedent to collect public funds for the support of a village high school, and allowed for many townships to follow their historic innovation. After the American Civil War, several landmark acts provided for further advancement and educational opportunity to a more extensive population. The First and Second Morrill Acts (1862, 1890) donated public lands to states and territories to provide agricultural and mechanical colleges. The Freedman's Bureau was established by the War Department in 1865, and promoted voting and education for refugees and freed slaves after emancipation, but the Jim Crow laws maintained much segregation between black and white students. It wasn't until the 1950s that Brown vs. The Board of Education in Topeka legally ended the segregation of schools (Anderson, 1988).

In the last 50 years, the American educational system has made great strides in an attempt to ensure a high quality free and appropriate public education for all students. In 1965, Congress passed the Elementary and Secondary School Act (Kantor, 1991), one of the most comprehensive pieces of legislation to date. The Act was a key piece of President Lyndon Johnson's War on Poverty, a domestic agenda designed to eliminate racial injustice and socioeconomic disadvantage by allowing for increased funding for education,

health care, and transportation. The original law contained six sections that were called “titles.” Since its original passage, the bill has been through numerous amendments and frequent reauthorizations all to ensure equality and improve the quality of public education. There have been several key amendments to the ESEA including the addition of aid to disabled children and bilingual education programs. Recent reauthorizations of the ESEA include the Education Consolidation and Improvement Act of 1981; Improving America’s Schools Act of 1994; and No Child Left Behind Act of 2001. In March of 2010, the Obama administration released its blueprint for the reauthorization of the ESEA.

According to the United States Department of Education:

The blueprint challenges the nation to embrace education standards that would put America on a path to global leadership. It provides incentives for states to adapt academic standards that prepare students to succeed in college and the workplace, and create accountability systems that measure student growth toward meeting the goal that all children graduate and succeed in college. (Vinovskis, 2009)

The current objectives of the Obama administration claim to be centered around providing every child with a complete and competitive education that will enable them to succeed in a global economy based on knowledge and innovation.

These objectives include:

- Higher standards and better assessments that will prepare students to succeed in college and the workplace.
- Ambitious efforts to recruit, prepare, develop, and advance effective teachers and principals, especially in the classrooms where they are most needed.

- Smarter data systems to measure student growth and success, and help educators improve instruction
- New attention and a national effort to turn around our lowest achieving schools.

While these objectives are lofty and theoretically advantageous, the legal maneuverings and logistics involved with implementing them with actual meaning and details for achievement have met with resistance. Innovative funding programs such as Race to the Top, the American Recovery and Reinvestment Act, and the American Jobs Act have succeeded in maintaining and furthering the advancement of educators, schools, and districts despite the economic recession of the last few years. However, because Congress has been unable to reach any type of agreement regarding the redesign and reauthorization of the ESEA, the Administration has provided individual states with flexibility within its parameters (Berry & Herrington, 2011).

### **The Purposes of Teacher Evaluation**

Teacher evaluation can serve a variety of purposes, and those are numerous and dependent. Haefele points out that a clear sense of purpose is necessary for the effective design of a quality teacher evaluation system. He explains that a system should, at minimum:

- Screen out unqualified persons from the selection and certification processes
- Provide constructive feedback to individual educators
- Recognize and help reinforce outstanding service
- Provide direction for staff development practices
- Provide evidence that will withstand professional and judicial scrutiny

- Aid institutions in terminating incompetent or unproductive personnel
- Unify teachers and administrators in their collective efforts to educate students (1993).

There are two purposes of personnel evaluation that are the most frequently cited: accountability and professional growth (see, for example, Danielson & McGreal, 2000, 2005; King, 2004; Peterson, 2002). Accountability as a purpose suggests the necessity of investigating competence of staff in order to guarantee that delivered instruction is both safe and effective (McGaghie, 1991).

Accountability has typically been interpreted as summative evaluation.

Performance improvement, conversely, suggests the need for professional development and growth of the teacher as an individual and as a practitioner.

Performance improvement is regarded as formative evaluation (Stronge, 2003).

Both purposes (growth and accountability) are necessary and desirable for evaluation to be most productive. Evaluation systems that include both dimensions are the most comprehensive and valuable, and the two should not be viewed as competing interests but rather as supporting and dual interests that are critical for improving educational outcomes. Moreover, the theoretical basis for the evaluation framework design should highlight the dynamic relationship between the teacher and the organization where the needs and goals of both support and fuse with one another (Stronge, 1995). A single evaluation system can be designed to meet both purposes successfully as long as the system is seen as a component of a larger mission. When the vision of the school and the furtherance of its goals are of primary importance, then a teacher evaluation program is tied to both teacher improvement and school improvement, and is

viewed as systemic rather than in isolation (McGreal, 1988). So an inclusive and thorough evaluation system will be designed with both purposes (formative and summative) and both levels (individual and organizational) in mind. Such a system should be accountability-oriented, adding to the evidence of individual goal attainment of each teacher as well as to the goals of the school and the educational institution as a whole. It should also be improvement-oriented, adding to the individual professional and personal growth needs of each teacher as well as the needs of the school (Stronge, Helm, & Tucker, 1995).

The ultimate mission of the educational process is to produce increased learning for all students, and the most important factor in the accomplishment of this mission is the performance and practice of the teacher. Therefore teacher evaluation can be considered a quality assurance mechanism (Kleinhenz and Ingvarson, 2004). An evaluation process, then, should be focused on educational efficiency and educational equity. The necessity of such a quality assurance method is emphasized by Stronge and Tucker (2003), when they report that the accountability purpose “reflects a commitment to the important professional goals of competence and quality performance”. They further explain that this accountability function is essentially the mechanism by which the effectiveness of educational services is judged.

The results of summative assessments also provide information that can guide the making of consequential decisions about the subject of the evaluation—the teacher. The justifications underlying most forms of evaluation are two-fold: either diagnostic information is needed, or evidence is needed for decision-

making. Teacher performance evaluation is not different (Avalos and Assael, 2006). When criteria is established and standardized, then comparisons between and among teachers become possible. The consequences of these comparisons can be useful for hiring and tenure opportunities, promotion decisions, or dismissals. Summative evaluation can also be used as a foundation for rewards and recognition systems that celebrate teachers who provide exceptional service. The status and image of the teaching profession has declined dramatically across the world, and teachers' often feel that their work is undervalued (Watt & Richardson, 2008). Evaluation data can address this concern by offering recognition for exemplary performance and demonstrated competence. These systems are necessary to make teaching an attractive career choice for students entering college and for retaining those teachers currently in the workforce that are effective.

### **History of Teacher Evaluation Practices**

Evaluation of teacher performance and practice is not a recent development. In fact, the ritual of teacher evaluation practices in America dates all the way back to the initial creation of educational institutions and schools in the early 18<sup>th</sup> century. There is early evidence that teacher evaluation began as a human resources/personnel function. In 1920, Nutt published his work *The Supervision of Instruction*, where he explains that supervision arose because of an amateur staff composed predominantly of women that was "the agency that will most adequately direct the work of all the teachers in the system, so as to improve the efficiency of individuals and to harmonize the work of the entire body

(4).” He also purports that professional training prior to entering a classroom teaching experience would be beneficial in that “there would be little need for any provision for ...supervision (5).” In 1923, Burton published a subsequent text entitled *Supervision and the Improvement of Teaching*. This work is the first modern iteration of what we now actualize as supervision that focuses on developing the practice of teachers (Domas & Tiedeman, 1950).

Classroom visits originated in the early 20<sup>th</sup> century, and gained popularity between 1910 and the 1930's. The characteristics of these early observations were substantially similar to those conducted for the next several decades. A supervisor would enter a classroom as unobtrusively as he could and would proceed to the rear of the classroom as to not detract from the lesson. He would then sit and take notes, and a “conference” meeting was held after the lesson where the teacher was first applauded for strengths and then presented with areas of weakness. In his text *Improving the Supervision of Instruction*, Spears explains that the evaluator was “first to commend the good but not to overstep the line that separates such commendation from flattery; otherwise the teacher would be in no position to accept the criticism to follow”. He was next to draw out the teacher as to whether the procedures followed in the classroom would reach the desired ends, and once the victim was “trapped” criticism was considered in order. (Spears, 1953, p.74) The teacher was then presented with the deficiencies identified by the supervisor, and ordered to implement ready-made procedures to correct them. The ritualistic nature of the “taken-for-granted procedure of

observation, judgment, and prescription done in a single visit” (Garman, 1986, p.150) had its origins in these early routines.

Even at the embryonic stage of principal evaluation, there was often skepticism as to the value of it. In 1936, Bernstein published an article in the New York Times entitled *Security of the Teacher in His Job*. In this article, he said: “The record of a teacher’s work, which is kept by the principal and which might be produced to show incompetence, is a weak instrument for the purpose. Standards vary greatly from school to school. One principal’s ‘satisfactory’ might be equivalent to another’s ‘unsatisfactory.’ School authorities, in recognition of this, are evolving a new system of rating” (Adams, 2006). This evolution continues today.

Scientific management and industrial psychology became influential during the industrial era with the work of Frederick Taylor and his counterparts. As a result, rating scales for teacher effectiveness were developed and implemented. The supporting logic is that if researchers could study the highly valuable teachers, then descriptors of their actions could be used to appraise and then develop those who were unsuccessful (Glanz, 2005). Standardized testing came to be during this same era, and teacher-rating instruments such as scorecards and checklists increased in widespread use with the idea that administrators could improve efficiency and save managerial time by eliminating the “lost motion in the teacher’s activity” (Spears, 1953, p. 66). These instruments were sometimes very lengthy, but not at all deep and focused on such aspects as the organization and appearance of the classroom and

individual student and teacher behaviors that could be easily measured and immediately corrected. The popularity of these instruments among administrators stemmed from objectivity, ease of use and efficiency, and the goal of improving instruction was noble; however, there was still a broad lack of consensus as to what characteristics actually comprised effective teaching.

While administrators felt they were useful, teachers regarded these instruments as critical, coercive, and in extreme cases “evil” (Spears, 1953, p. 75). The visits were viewed as having no real purpose or goal of professional improvement, and instead were merely faultfinding missions. Teachers reacted with dissatisfaction, and policymakers in the field began to construct more representative and sensitive evaluation measures. The articulated purpose of supervision became “professional growth” instead of criticism, and democratic tenets were adopted to counter the teachers’ frustration.

Since that time, teacher evaluation has been conducted as a process whereby an administrator (principal or supervisor) observes a classroom teacher performance a limited number of times and bases recommendations for improvements on this limited number of contacts. Morris Cogan, Robert Anderson, and Robert Goldhammer at Harvard heavily influenced this model of evaluation in the 1950’s and 1960’s with their Clinical Supervision model, and little has been done to alter it to date (Wiedmer, 1995). Goldhammer proposed a five-step process to clinical supervision (1969):

1. Preobservation Conference
2. Observation
3. Analysis and Strategy

#### 4. Postobservation Conference

#### 5. Postconference Analysis

This process is credited with egalitarianism; however, the actuality of it is seldom viewed as useful or meaningful in any way. Ponticell and Zepeda (2004) illustrate that the classroom observation visit is both ritualistic in purpose and direction.

One teacher in their analysis of narrative interviews of practitioners said:

“The principal starts and ends with a list. The pre-observation conference consists of his handing me an observation list and telling me that as long as my lesson covers all the elements delineated, I will be okay. In the post-observation conference, he points out all the errors I made on the list and indicates how many points I lost. He directs me to sign the form: ‘Fix these for your next observation.’”

Vestiges of this ritual are still seen in schools today. There have been many modifications of the process, but the skeletal framework is still the same and remains for both teachers and principals a “tiresome chore, one that takes an enormous chunk of time” (Black, 2003, p.38 as reported by Hazi & Rucinski, 2005).

### **The Evaluation of Teaching Is Essential to Effective Education**

All stakeholders- parents, practitioners, and policymakers- agree that education in America is grounded in placing highly skilled and effective teachers in all classrooms. Education as an institution cannot improve without a workforce that is measurably competent; however, the nation (and the world) has been unable to formulate a reasoned consensus as to which measures demonstrate competence, and thus a practical set of standards and assessments has never been achieved.

Several approaches of measuring teaching effectiveness have been proposed for licensing and certification of new teachers as well as similar approaches for evaluation and continued employment of existing teaching staff. *A Nation Prepared: Teachers for the 21<sup>st</sup> Century* (Force, 1986) was a landmark report that concluded what stakeholders already knew: high quality teachers are a critical element in the improvement of student outcomes. Efforts to improve teacher evaluation methods have resulted from this and other reports of the mid '80s that unnerved the education community. For more than three decades, policymakers have undertaken many and varied reforms to improve schools. These reforms range from new curricula to new standards and tests to new evaluations and governance models. One lesson continues to emerge as a recurring theme of these efforts: the “teachers are the fulcrum determining whether any school initiative tips toward success or failure” (Darling-Hammond, 2010). There is no question that every facet of educational improvement depends on high-quality teachers for its success. There aren’t many areas of agreement among teachers, lawmakers, and the general public, but improving teacher quality as a direct and promising strategy for improving public education outcomes is one. There is universal concurrence that teaching matters. “Without capable, high quality teachers in America’s classrooms, no educational reform effort can possibly succeed” (Stronge & Tucker, 2003, p.3).

It is quite an elementary idea that there is a connection between teaching and learning, and that the connection is best realized when we have high quality and effective people working with all students at all times. Unfortunately, the

definition of “high quality ” has proven difficult to pinpoint, and there are a myriad of ways that “effective” is defined (Cruickshank & Haefele, 2001). Teachers drive the learning process, and it is absolutely critical that we have the best teachers in that driver’s seat.

The nature of evaluation, both formative and summative, must be rooted in organizational effectiveness as well as teacher quality. The formative focus of evaluation is best implemented through sustained, meaningful feedback from a variety of sources. These sources provide evidence to first document the effectiveness of teaching performance, and then using this evidence as data to assist professionals in improving that performance. When this first focus is actualized, then the second focus of evaluation- the summative piece- becomes less frightening. Teacher quality and accountability become foregone conclusions when the development and growth has been consistent and ongoing.

### **Traditional Teacher Evaluation**

Teacher evaluation has historically been a highly controversial subject. There is both mixed empirical evidence about its impacts on learning and conflicts of interest between key stakeholders within the systems. Consequentially, the entire whole of evaluation has often been a meaningless enterprise deriving its origins from requisite bureaucratic ritual exercises in schools, and generally being tolerated by both the teachers themselves as well as the evaluators (Danielson, 2001: Holland, 2005; Marshall, 2005).

In most countries including the United States, administrative personnel within the school building- generally a principal- conduct the classroom

observations of teachers. Several researchers have criticized the longstanding practice of holding yearly scheduled evaluations because they fail to provide an authentic snapshot of day-in, day-out teaching performance. Researchers also mention the lack of constructive feedback, coaching, or suggestions for improvement as disadvantages (Klinger et al., 2008; Daley and Kim, 2010; Danielson, 2011; Marshall, 2012; Papay, 2012).

Charlotte Danielson is one of the premier experts in the area of teacher evaluation, and her work is widely respected and adopted. In her book *Teacher Evaluation to Enhance Professional Practice* (2005), she identifies six primary areas of deficiency in traditional evaluation practice. They are:

- Outdated, limited evaluation criteria
- Few shared values and assumptions about good teaching
- Lack of precision in evaluating performance
- Hierarchical, one-way communication
- No differentiation between novice and experienced practitioners
- Limited administrator expertise

Danielson explains that the results of these deficiencies, in isolation or in combination, have disastrous consequences to the quality of education. They lead to a culture of “passivity and protection” where skepticism, dishonesty, and suspicion abound. Teachers do not feel comfortable having a frank and open conversation with administrators regarding difficulties they may be experiencing for fear of retaliation in the evaluation. This environment encourages stagnation and discourages any informed risk-taking in instructional pedagogy. Teachers may not respect the credibility of the principal and his expertise in performing the observation, nor do they anticipate any value in learning and growth through the

“dance” of the evaluation process. As a result, many teachers perform “canned lessons” that demonstrate all of the measurable behaviors on the observation checklist, in some cases replicating the exact same lesson over a period of years and coaching students on how to behave and respond. And they receive superior ratings. No professional value results from the futility of this exercise, and neither the administrator nor the teacher learns anything that he/she didn’t already know. The entire process is meaningless, unrewarding, tedious, and time-consuming.

Those experiences, while empty, are still on the positive side of the continuum of traditional evaluation systems. Much worse are the schools where there is a culture of “gotcha” and principals use the evaluation process to assign low ratings with no justification or evidence of such just to dismiss those people they may not like. In these schools, the process moves from merely meaningless to punitive and damaging (Danielson, 2005).

### **Current Status & Recent Developments in Teacher Evaluation in America**

There are generally two main goals associated with teacher evaluation systems. The first goal is to improve teachers’ practice by providing performance feedback, and the second is to identify, remediate, and in extreme cases dismiss those teachers that perform below acceptable levels. While these goals are well intentioned, they are seldom realized. Most teachers believe that evaluation systems do very little to improve their practice. A report entitled “The Widget Effect” published by the New Teacher Project, in June of 2009, thoroughly investigated this pervasive belief:

When it comes to measuring instructional performance, current policies and systems overlook significant differences between teachers. There is

little or no differentiation of excellent teaching from good, good from fair, or fair from poor. This is the Widget Effect: a tendency to treat all teachers as roughly interchangeable, even when their teaching is quite variable. Consequently, teachers are not developed as professionals with individual strengths and capabilities, and poor performance is rarely identified or addressed. (Weisberg et al, 2009)

This report indicated several key indicators that have led to what has been termed the “systematic de-professionalization of teachers.” These characteristics include:

- Performance evaluations are perfunctory and infrequent
- Primary use of evaluations is to identify incompetence
- Teachers expect to receive the highest possible rating, even during their first years in the classroom
- Evaluations do not yield meaningful feedback for teachers, and professional development is not aligned to evaluations
- Administrators are poorly trained to evaluate and districts do not prioritize the process
- Teachers who receive feedback for improvement during the evaluation process tend to feel singled out, often unfairly

In a recent (Aldeman, 2011) study conducted by Education Sector, only 26 percent of teachers labeled their current evaluation systems as “useful and effective.” Both teachers and administrators consistently relate that poor teacher performance is pervasive in American schools; despite this fact, less than 1 percent of teachers are identified as “unsatisfactory” in performance evaluations. The byproducts of the inflated evaluation system are many, and a stark and unfair one is that true excellence cannot be identified when ratings in the excellent category are the rule rather than the exception. Conversely, teachers

with poor performance are not aware that anything about their teaching practice needs improvement.

Other findings in the report indicate that there is a glaring lack of specific feedback to inform the professional development process. 3 out of 4 teachers in the study reported receiving no specific performance feedback for improvement at all during their last evaluation cycle (Weisman et al, 2009). Without feedback, evaluation is both uneven and inadequate, and amounts to an accountability exercise rather than a resource and tool for improvement.

### **Current Status & Recent Developments in Teacher Evaluation Internationally**

Outcry for instructional quality has led many countries to establish one form of teaching performance assessment or another. The Organisation for Economic Cooperation and Development, OECD, is a collaborative international group founded in 1961 that's mission is to "promote policies that will improve the economic and social well-being of people around the world." Its Directorate of Education conducted a massive global *Review on Evaluation and Assessment Frameworks for Improving School Outcomes* in 2009, and followed this review in 2013 with a follow up report entitled ***Teachers for the 21st Century: Using Evaluation to improve teaching***. Those two, along with the OECD's *Teachers Matter: Attracting, Developing and Retaining Effective Teachers* that was published in 2005 are three examinations of the current academic and policy research on teacher evaluation in primary and secondary education on a global scale. The following section is based on these three reports and supplemental papers derived from their data.

In the *Review on Evaluation and Assessment Frameworks for Improving School Outcomes*, 26 school systems across 25 different countries participated in the review. The overarching policy question was “How can assessment and evaluation policies work together more effectively to improve student outcomes in primary and secondary schools?” The Review additionally focused on five key issues for analysis:

- Designing a systemic framework for evaluation and assessment
- Ensuring the effectiveness of evaluation and assessment procedures
- Developing competencies for evaluation and for using feedback
- Making the best use of evaluation results
- Implementing evaluation and assessment policies

In recent years, some countries have demonstrated a growing desire to develop evaluation systems as a critical piece of broader teacher and school improvement policies (Peterson, 2006). Existing schemata of teacher evaluation in OECD countries assume multiple formats. Educational context and tradition, purposes of evaluation emphasis, and the actors involved in the implementation differ largely from one country to another, and the scope and methods of teacher evaluation, criteria and standards used, and data gathering instruments differ largely as a result. There are also widely diverse consequences of the evaluation processes for teacher careers. Single salary schedules and single promotion tables remain prevalent; however, several countries have recently attempted to link the teacher appraisal system to either professional development opportunities or to a system of recognition and rewards, whether financial or not.

The recruitment and retention of high quality prospects to the teacher profession is an international concern in the world of education. According to the Australian Department of Education, people who generally are motivated by intrinsic rewards tend to choose teaching as a vocation; however, extrinsic rewards such as remuneration are the greatest factors that influence people not to choose a career in teaching, and the greatest factor that leads those who are successful to exit the profession (Lavy, 2007). Also, the single salary scale is predominant, but a few European countries do offer a pay increase upward from that predetermined base for exemplary performance. Romania, for example, has established a mechanism in which the most successful teachers are offered the opportunity to compete for a salary raise (albeit temporary) from 15% during a year to 20% during four years (Dolton & Marcenaro-Gutierrez, 2011).

There is extensive variation in teacher evaluation and appraisal systems across the world. Countries like Korea, Poland, Australia, Sweden and New Zealand have national systems that are highly sophisticated and monitored for implementation at the national level, while Denmark, Iceland, Spain, the French Community of Belgium, Finland, and Norway have not adopted national or state framework policies and are instead monitored locally or at the individual school level. There is also significant variation as to the point in time at which a teacher enters the systematic appraisal process, and the uses of the data. 13 countries including Canada, the United Kingdom, Ireland and Italy have a probationary period for teachers that may range from 6 months up to 5 years. These teachers are observed more frequently than regular teachers, and data from the appraisal

system for them are used for certification or registration into in-service teaching. 17 countries also implement evaluation systems for regular performance appraisal, and 5 countries used the data for promotion decisions. Only 3 countries (Chile, Korea, and Mexico) use the data from performance evaluation as a component of a reward scheme. These appraisals are not mandatory, and teachers must voluntarily apply for the privilege (Isore, 2009).

Teacher evaluation at the international level is still very firmly rooted in the classroom observation ceremony. There are also a wide variety of other inputs used, such as teacher self-appraisals, teacher portfolios, teacher content knowledge testing, and of course, student outcome measures.

### **The Value Added Model**

Value-added scores are estimates that use statistical methods to predict the effect of individual teachers, classrooms, and schools by adjusting for the prior achievement of students and certain measured factors (Kane & Staiger, 2012). Most authors are not convinced that current models of value-added calculations are valid and reliable enough to be used confidently for evaluation of the individual performance of a teacher, and that they offer significant statistical, methodological, and sampling challenges (Kupermintz, 2003; Braun, 2005; Goe, 2008). There is also a very great risk in isolating student achievement in terms of performance on one value-added measure, and ignoring the innumerable other factors that may influence student achievement independent of the teachers' impact- attendance, health, socioeconomic status, family background, classroom culture, etc. Such shortsighted calculations can have disastrous results, including

the realistic possibility that teachers will be punished or rewarded for scores that are beyond their actual control (Kane & Staiger, 2012; Kupermintz, 2002; McCaffrey et al., 2003; Ingvarson, Kleinhenz and Wilkinson, 2007). Policymakers have responded, and in April of 2008 the New York State legislature made the bold decision to ban the use of any standardized test scores for the purposes of teacher evaluation.

Alternatively, proponents of the Value-Added Model in Florida have linked salary/bonus mechanisms to high value-added measures of student learning. Their “Special Teachers are Rewarded” (STAR) model is one of several merit based, pay-for-performance schemes that are developing around the country (Buddin, et al, 2007). The Obama administration supports such pay-for-performance models, and gave special consideration to grant applicants who included the use of student test scores in teacher evaluation. This rationale is supported by other research that suggests that there is scant evidence to support that value-added measures are so grossly biased as to be “directionally misleading”. Contrarily, Kane and Staiger (2008) assigned a small sample of teachers to specific rosters by lottery, and were unable to reject a non-bias. They concluded that the value-added measures statistically approximated teacher effects that were “causal” for student achievement. Another study was conducted to test student class size, but in its design it also randomly assigned students to teachers. This study was reanalyzed in 2009. The reanalysis reported that effects on student achievement of individual teachers were actually larger than many of

the same results that were reported in the value-added investigation (Nye, Konstantopoulos and Hedges, 2004).

Research seems to support that value-added measures do indeed seem to provide some information about a teacher's individual impact; However, the value-added measures themselves may need to be scaled up or down if evidence of bias at the end of the year is suspected or reported. Value-added measures can have meaning, but should only be used as one component of a larger composite evaluation of a teachers' performance (Kane & Staiger, 2009).

The most recent and perhaps interesting of value-added research was released in the culminating report from the MET project. In the final stages of analysis, the researchers were able to use actual data gathered from the first two years of the study on value-added measures of teachers, and were able employ the full power of random assignment in year three to determine if in fact these results were valid. From the report:

By randomly assigning students to teachers, we made it very unlikely that the students assigned to seemingly more or less effective teachers would be different in measured or unmeasured ways. Therefore, following random assignment, we studied the achievement of students assigned to teachers with differing prior measures of effectiveness. Looking across all the sample teachers and their predicted outcomes, we asked two questions:

- (1) Did the measures of effective teaching successfully identify sets of teachers who produced higher student achievement gains on average?
- And (2) did the magnitude of the differences correspond with what we would have predicted based on their measured effectiveness in 2009–10?

The results of the study supported the contention that identification of teachers that produce higher student gains is indeed possible. The effectiveness measures from year one did identify the teachers that produced higher mean student achievement gains following the power of random assignment.

Collectively, the teachers labeled as highly effective registered greater gains in student achievement than their colleagues in the same subject, grade, and school. A second finding was that the level of the achievement gains these teachers generated was consistent in magnitude with expectations of researchers. The measures of effectiveness were gathered prior to random assignment under conditions of the usual manner of assignment of students to classrooms; however, the measures were able to predict a teacher's impact on students that was statistically consistent when the students were later randomly assigned (Kane Staiger, 2013).

### **The Importance of Feedback, Trust and Follow-Up**

Feedback is defined as *information that provides the performer with direct, usable insights into current performance*. The meaningful application of feedback will be based on observable and tangible discrepancies between the current state of performance and the intended best state of performance (Wiggins, 1998). In order for feedback to be valuable and to steer improvement, it must adhere to the eight principles described by Gilbert in *Human Competence* (2007). In this book, he describes the information designs that maximize support for performance. One of them is to “provide frequent and unequivocal feedback about how well each person is performing.” He adds that this feedback should be communicated within a framework of comparison with an exemplary standard. It is only within this construct that mastery of complex skills such as teaching can be achieved. Mastery is solving complex problems by reacting to the feedback received within the actual circumstances, situation, and setting (Wiggins, 1993).

Feedback, when it is accurate and trustworthy, is holistic information. The ability and the flexibility to consume and understand critical feedback from all levels in a useful and productive way is a necessary condition for the exercise to be beneficial (Nowack, 1992, 2009). In order for this condition to be met, the recipient of the feedback needs to both trust and value the feedback provider.

Rousseau et al. (1998) conclude that “trust, as the willingness to be vulnerable under conditions of risk and interdependence, is a psychological state that researchers in various disciplines interpret in terms of ‘perceived probabilities’, ‘confidence’, and ‘positive expectations’- all variations on the same theme.” A worker’s trust in the manager- in this example, a teacher’s trust in the principal- can be divided into two major categories: task-oriented and relationship oriented. Research in leadership theory has investigated these two variables for more than half a century, and includes such landmark studies as the Ohio Leadership Studies (Korman, 1966) and The Managerial Grid (Blake and Moulton, 1968). Task-oriented feedback is rooted in the performance of the job itself. Relationship-oriented feedback is rooted in providing support (both personal and socio-emotional) and maintaining open lines of communication.

In order to trust in the feedback of the principal, a teacher must consider him competent to offer the feedback. If not, it is unlikely that the teacher will support the results, which renders them meaningless (Bass & Stogdill, 1990). The definition of competence is *requisite or adequate ability or quality*. Literature is consistent in the conceptualization of competence as whether or not the evaluator/manager has the capability to perform the assigned tasks, and

researchers have emphasized the perception of competence as an “essential antecedent” of trust in the results.

### **360 Degree Feedback Models in other Professions**

Private industry, for-profit institutions, and some government organizations, including the military, have already implemented and embraced 360 degree feedback mechanisms and the importance of multi-rater assessments. The most current literature in the study of both leadership assessment and performance evaluation asserts that 360-degree assessments are a very effective method when implemented with fidelity. There are many different models and structures of the multi-rater or 360-degree assessment, and they include a wide variety of formats and overlap in content. Senior executives, managers, entry level employees, and customers/clients are asked to provide their feedback to the evaluation subject, and that feedback is then collected and presented to the subject in a report that explains the results. Much like the instruments themselves, the reports come in a wide array of formats which range from individual item analysis to comparative results within groups and organizations (i.e. percentile groups). Specific professional competencies are identified and assessed at each level, and can include categories such as leadership, effectiveness, and relatability.

Research also suggests that 360-degree assessments can only be effective in conjunction with organizational support and will not be useful in isolation. In order to establish organizational support and ensure that usable, meaningful data is collected, the following best practices should be adhered to:

- Clarify the purpose for every member of the organization
- Clarify the rater anonymity, accountability, and selection
- Prepare participants
- Review and interpret feedback results
- Develop an action plan
- Follow-up

Dependence upon only traditional talent metrics and performance

evaluation data is insufficient to improve effectiveness. Research says that an alternative approach should highlight the necessity of garnering feedback from many levels: peers, coaches, mentors, and customers. This multi-faceted approach to feedback can uncover unique traits and identify hidden talents. Even word-of-mouth information is valuable to managers when it is credible. Such evaluation data can identify “intangibles” such as the level to which and employee is influential, respected, and connected to the organizational network. It can also be used to pinpoint talents or shortfalls that require improvement (Schurr & Tambe, 2008). Each piece of information that is gathered from a different perspective adds nuance to the evaluation procedure and helps to shape a more complete dimensional picture of performance.

### **Student Feedback**

The use of student feedback as a reflection of teacher performance is viewed as important for evaluation. More and more authors suggest using this form of data. There is already a rather extensive body of literature that investigates using student ratings at the university level, and a similar body is growing on the topic for K-12 education. Cashin (1995) mentions an excess of 1500 references reporting student evaluation of quality instruction from 1971

through 2000. These readings illustrate the controversial nature of the subject of student feedback, particularly at the primary and elementary levels. The controversy itself gives support to the importance of the subject, and the serious dialogue surrounding the topic asserts such (Wilkerson et al, 2000). Teachers across the nation have expressed disapproval with current practices of evaluation, and while most educators appreciate the idea of student feedback as a component of evaluation, concerns about student bias consistently emerge. These concerns are relative to the discrimination abilities of students to provide accurate commentary on instruction, as well as worries about maturity and prejudice (Shepherd & Trank, 1989; Vollmer & Creek, 1989). Still, teachers collectively feel that the students can provide a more accurate and representative picture of classroom life than current evaluation methods can, and that their perceptions are more meaningful.

Proponents of using student feedback as a component of teaching evaluation provide indications from theory and from application that children in K-12 environments have the skills and knowledge to offer valuable insights into the strengths of classroom instruction as well as the areas of potential improvement (Wilkerson et al, 2000). Individual student bias can absolutely be projected to influence the judgment of a respondent; however, aggregating the student feedback into a pool of the entire class helps to decrease the subjective distortion, and creates an accurate view of teacher behavior and its effect on student engagement (Joyce & Peck, 2005). These ratings are quite valid, and

can be used to provide educators with valuable information to inform professional growth and development.

Additionally, in traditional schools, the students are the only people who consistently observe the performance of any particular teacher on the job each day. This conclusion is supported by Omotani (1996) and Peck (1998), both of whom report that students in typical learning environments are the “ideal contributors” of data on teacher quality for the entire instructional period because they are the only ones who regularly see a teacher at work. Students of any grade level and any age have the ability to provide a fair assessment of teacher evaluation (Savage and McCord, 1986).

The culminating findings from the MET Project (Kane & Staiger, 2013) support the use of student feedback for teachers’ appraisal. The initial 2010 report finds that a well-constructed survey of student perceptions can provide consistent feedback data on aspects of teaching performance that are predictors of student achievement (Kane & Staiger, 2010).

Student feedback surveys can present valuable data regarding the perceptions that students, parents and other stakeholders who have ongoing, consistent interaction with a given teacher interpret that teacher’s quality of craft (Peterson et al., 2006; 2003; Jacob and Lefgren, 2005). Despite their value, such instruments are seldom implemented as components in formal teacher evaluation systems.

## **Teacher Self-Efficacy**

The study of self-efficacy in teacher performance is also not a recent development. Such information grew out of widely known psychological research of Bandura's social cognitive theory and his revised descriptions of self-efficacy illustrated in his seminal 1977 article, "Self-efficacy: Toward a unifying theory of behavioral change". Bandura defines self-efficacy as "a judgment of one's capability to accomplish a given level of performance" (1986), and then later as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (1997). More recently, the idea of self-efficacy has taken a more emphatic role in educational research in terms of both student self-efficacy and teacher self-efficacy. According to theory, self-efficacy refers to the "future-oriented belief about the level of competence a person expects he or she will display in a given situation" (Tschannen-Moran & Hoy, 2001). These beliefs influence several psychological factors, including thought patterns and belief systems, as well as emotional reactions that stimulate actions where people invest significant energy in the attainment of goals, persistence in the face of obstacles, coping with minor setbacks, and exert some degree of control over situations that have an effect on their lives (Bandura, 1986, 1993, 1997). Teachers with high self-efficacy have not only a desire to succeed, rather an expectation that they will succeed in teaching students and managing them effectively. Moreover, this expectation influences their interpretation of successes and disappointments, the approaches they employ to cope with challenges, and the standards they set for themselves (Bandura, 1997; Ross, 1992). The

benefits of strong self-efficacy beliefs have been demonstrated through research: less stress and burnout among teachers and increased job satisfaction, and these have been linked to instructional practices and student achievement (Ashton and Webb, 1986; Ross, 1998).

The classroom implications of reports of teacher self-efficacy are numerous. There are indicators within the constructs of teacher efficacy research that show aspects of increased productivity and effective teacher behavior and the resulting positive impacts on school culture and organizational effectiveness. Teachers with a high level of self-efficacy “may be more likely to adapt to and moderate dynamics in schools whose students come from different backgrounds or present particular challenges” (Isore, 2009). According to TALIS data, the appraisal and feedback that is given to a teacher can influence his/her feelings of self-efficacy and teachers with high levels of self-efficacy are associated with several other positive practices and behaviors that influence the effectiveness of their performance and the quality of student learning in their classroom.

There are also challenges associated with measuring teacher efficacy. Bandura’s social cognitive theory introduced two separate factors involved within the definition of self-efficacy: personal efficacy, and outcome expectancy. Researchers agree that teachers’ personal efficacy is a measure of the individual’s personal sense of competence as a teacher. It has been much more difficult to arrive at a meaning for the second dimension of Bandura’s theory for teachers since the various elements of teacher outcomes are widely distributed. Several different instruments to measure self-efficacy have gone through many

iterations in the course of four decades, and drew from many researchers including Emmer and Hickman (1990), Riggs and Enochs (1990), Gibson and Dembo (1984), Soodak and Podell (1996), and Guskey and Passaro (1994).

The Teachers' Sense of Efficacy Scale was developed at the Ohio State University by a team of researchers interested in solving this dilemma. They expanded the work of Bandura and many other researchers working with cognitive theory on self-efficacy to create the instrument used in this study. Validity and reliability data on the TSES are offered in the Methodology Section.

An additional value of a strong teacher evaluation system is that formative appraisal has been shown through research to increase teachers' self-efficacy. In that vein, the inclusion of efficacy measures to an effective evaluation system may have mutually beneficial effects on both (OECD, 2013).

### **The MET Project**

In 2009, the Bill and Melinda Gates Foundation funded a project that was designed to improve the quality of data about teaching effectiveness. The project is called the Measures of Effective Teaching (MET), and its goal was to rigorously develop and test several measures of teaching effectiveness beyond the traditional test scores and principal evaluations. More than a dozen reputable academic, non-profit, and for-profit agencies crunched data collected during two consecutive school years (2009-10 and 2010-11). These data include over 3000 classrooms of teacher volunteers across several educational agencies in the United States, and some are among the largest school districts in the country.

There are five critical research areas investigated within the MET Project as measures of effectiveness:

- Student achievement gains on state standardized assessments and supplemental assessments designed to measure higher-order conceptual thinking
- Classroom observations and teacher reflections
- Teachers' pedagogical content knowledge
- Student perceptions of the classroom instructional environment
- Teachers' perceptions of working conditions and instructional support at their schools

The MET Project was a massive study conducted over 4 years, and a great many reports have been generated as a result. The final conclusions and findings were released in 2013, and contained some very powerful determinants.

The key findings from their final analysis report that impact the current study include:

- 1.) Effective teaching can be measured- as explained in the Value-Added section of this chapter, the MET project was able to determine that through measures adjusted for external factors, and by using the power of random assignment, the data demonstrated that groups of teachers who are more effective can statistically be identified. Further, the level of the academic growth that more effective teachers achieve was consistent with predicted values.
- 2.) Balanced weights indicate multiple aspects of effective teaching- while student achievement is an important measure, the MET project found that multiple measures of teacher performance produce more reliable and consistent ratings than value-added measures alone. "Estimates of teachers' effectiveness are more stable from year to year when they combine classroom observations, student surveys, and measures of

student achievement gains than when they are based solely on the latter” (MET, 2013).

- 3.) Surveys that measure student perceptions and classroom observations are able to provide meaningful, valuable feedback to teachers- These can also enable district leaders to prioritize funding decisions in staff development to focus on the largest disconnects between teachers’ observed actual performance and the established standards for effective practice.

(Kane et al, 2008, 2009, 2012, 2013).

## **CHAPTER 3: METHODOLOGY**

### **Research Design**

The primary purpose of this study was to determine if relationships exist between student achievement (as measured by value added test scores) and teacher self-efficacy, student perceptions of teacher effectiveness, and principal observations of teachers in K-12 public schools in the southeastern United States. In this chapter, the methods used by the researcher are described. The study was designed as a correlational study using survey and observational methodology.

Based on previous research findings and other conceptual evidence from the review of related literature, the following objectives were written in the form of research hypotheses to be tested:

1. Teachers working in an elementary, middle, and high (K-12) public school in the southeastern region of the United States with higher measures of self-efficacy will have higher value added test scores.
2. Teachers working in an elementary, middle, and high public school in the southeastern region of the United States with higher principal observation scores will have higher value added test scores.
3. Teachers working in an elementary, middle, and high public school in the southeastern region of the United States with higher scores in the students' perceptions of teacher effectiveness will have higher value added test scores.

4. The students' perceptions of teacher effectiveness and the measures of self-efficacy will be more closely correlated to the value added test scores of the teacher than the principal observation.

The dependent variable is the calculated Value Added Measure of each individual teacher. The independent variables are the self-efficacy scores, principal observation scores, and student perceptions scores of each teacher as measure by the instruments described below.

Approval to conduct this study was obtained from the Louisiana State University Institutional Review Board (IRB# E8198) and the case study school, Knightsville Academy. The following sections will discuss the population/sample, the instrumentation, the data collection methods, and the proposed analyses.

### **Population and Sample**

The target population for this study is defined as elementary, middle, and high school teachers in public schools in the southeastern region of the United States. The case study school for this data collection will be referred to as Knightsville Academy. The school was selected because it spans every grade from Kindergarten through twelfth grade, and the enrollment is based on a lottery system which results in a population that is statistically designed to be evenly distributed (25% white boys, 25% white girls, 25% minority boys, 25% minority girls), thus eliminated the extraneous variables of race and gender. There are no academic criteria for admission to Knightsville, but a minimum 2.5 grade point average is required to remain in attendance there. It is a public school that draws from the rural population of a parish in south Louisiana.

On March 8<sup>th</sup>, 2013 the researcher presented the purpose, rationale, objectives, and design of the study to the faculty of the Knightsville at the invitation of the Principal. The initial study design dictated that a random sample of the faculty/classrooms would be used as research subjects the size of which would be determined using Chocran's formula; however, 100% of the faculty enthusiastically volunteered to participate in the study, and so a census method of data collection was selected and informed consent was acquired from all subjects (Appendix A), along with permission to obtain the measures of confidential data that are critical to the study.

The researcher obtained a list of all of the K-12 teachers at Knightsville from the principal of the school. There were 68 teachers on the list. Four individuals were removed from the list because they serve as Master Teachers or Instructional Coaches, and therefore do not receive Value Added Measures. The 64 remaining teachers all participated in the case study, and we had a 100% response rate.

### **Instrumentation and Data Collection**

Data collection happened in five phases: self-efficacy survey administration, demographic data, student perceptions survey administration, principal observation scores collection, and Value Added Measures collection.

#### **Phase 1: Self-Efficacy Survey Administration**

The Teacher Sense of Efficacy Scale (Woolfolk-Hoy) was used to measure the kinds of factors which create difficulties for teachers in their school

activities and their perceived measure of impact upon those factors. Permission to use the instrument was granted by the developer (Appendix B).

The instrument (Appendix C) consists of 24 items evaluated on the same question: “How much can you do?” The answers are weighted on a 9 point anchored scale where a score of 1 is “nothing” and a score of 9 is “a great deal.”

The validities and reliabilities of the instrument have been established in many previous studies. In prior studies, factor analyses were conducted to determine how participants responded to the questions. Three moderately correlated factors have been consistently found: Efficacy in Student Engagement, Efficacy in Instructional Practices, and Efficacy in Classroom Management. To determine subscale scores for each of these factors, the unweighted means of the items that loaded on each factor were computed.

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

Table 3.1 Reliabilities of the Teachers Sense of Efficacy Scale

	Mean	Standard Deviation	Alpha
<b>TSES (whole)</b>	7.1	.94	.94
<b>Engagement</b>	7.3	1.1	.87
<b>Instruction</b>	7.3	1.1	.91
<b>Management</b>	6.7	1.1	.90

These groupings were:

*Efficacy in Student Engagement:* Items 1, 2, 4, 6, 9, 12, 14, and 22

*Efficacy in Instructional Strategies:* Items 7, 10, 11, 17, 18, 20, 23, and 24

*Efficacy in Classroom Management:* Items 3, 5, 8, 13, 15, 16, 19, and 21

On March 8<sup>th</sup>, the majority of participating teachers attended the presentation of the purpose, rationale, objectives, and design of the study to the faculty of Knightsville Academy. 90% of the teachers completed the Teachers Sense of Efficacy Scale during this meeting. The researcher contacted each of the teachers that were absent from this meeting during phase 3 of data collection, and the remaining scores were obtained then. Completed surveys from 68 teachers were received, and the results are reported in figure \_\_\_\_\_. Data from the 4 master teachers and instructional coaches were removed from the final analysis.

## **Phase 2: Demographic Data Collection**

The researcher obtained a complete faculty email list from the principal of Knightsville Academy. On March 22, 2013 the Knightsville Teacher Demographic Webform (<http://www.cain.lsu.edu/MSAwest>) was delivered via this medium to the research subjects to collect demographic information related to the following:

- Years experience
- Grades taught
- Subjects taught
- Gender
- Race
- Age

These demographic variables were selected based on previous studies collected in educational research, and were drawn from literature on teacher evaluation. Individual reminders were sent to all non-completers at 48-hour intervals. Complete demographic data was collected by 100% of participants by the implementation of phase 3. Each teacher was then assigned a unique 3-digit identifier to ensure anonymity and confidentiality.

### **Phase 3: Student Survey Perceptions Administration**

The four rating instruments used in this phase of data collection consisted of 20 items that are positive descriptors of teacher behavior and are a part of the 360-degree system developed as a component of the School Improvement Model (SIM) project at Iowa State University. The researcher contacted the developers of the instruments and obtained permission to use them in the study (Appendix D).

In coordination with the administration of Knightsville Academy, the week of April 29 through May 3, 2013 was designated for the administration of student perceptions of teacher effectiveness surveys. The researcher traveled to the school site, and spent the week there overseeing the study. The researcher developed scripts (Appendix E) to be read by each administering teacher assuring the students that the data would be anonymous and that the teachers would not receive the results until the following school year. Additionally, the researcher paired and swapped teachers in the same grade levels so that no teacher would survey his or her own students. This was done due to the sensitive nature of the data and to ensure students trusted that the information was

anonymous and would never be used in a punitive manner. Completed packets were returned by the administering teacher. This phase resulted in over 2000 data points. The contents of each teacher packet were checked for proper coding and clean marking by the researcher.

The student perceptions instruments utilized in this study were originally developed in 1990 and 1991 as part of research conducted by Omotani (1992) and Weber (1992) in studies investigating the discrimination power of survey questions on feedback measures. The instruments are categorized into respective levels: lower elementary (K-2), upper elementary (3-5), middle school (6-8), and high school (9-12). (Appendix F). These surveys were originally created by the School Improvement Model (SIM) in their efforts to include student evaluations of teacher performance as a component of a teacher evaluation system. Many of the survey items selected for inclusion in the student feedback forms were chosen from a pool of valid, reliable, and legally discriminating questions identified by Judkins (1987). The rest of the items were developed by teachers. The items and surveys went through several stages of revision prior to the adoption of the final 1991 versions.

The student feedback instruments each are composed of 20 items. Each item was designed to be read as positive descriptors of teacher behavior. Omotani (1992) and Weber (1992) research found the instruments to be valid, reliable, and discriminating. Prior to their use in the current study, each instrument had also been tested for grade-level readability. The K-2 survey uses a three point scale with the following values: No=0, Sometimes=2, and Almost

Always=4. The three other surveys (3-5, 6-8, and 9-12) use five-point, Likert-type scales with values of: Never=0, Not Often=1, Sometimes=2, Usually=3, Almost Always=4. This type of scale means that for each instrument regardless of grade level, teachers receiving the Almost Always rating for all of the 20 items would get a total score of 80. A mean score between 1-80 for each teacher was calculated based on all of the surveys collected from that teacher's students.

The Louisiana State University Evaluation and Assessment Center designed the instruments for coding (Appendix G), and assisted in processing the data and creating files to use with the SPSS statistical software package.

#### **Phase 4: Principal Observation Scores Data Collection**

The observation measures were conducted using the TAP Rubric, the district mandated evaluation form (Appendix H ). The TAP Instructional Rubrics "utilize a 5-point rating scale, with a 5 indicating Exemplary, a 3 indicating Proficient, and a 1 indicating Needs Improvement." (LaDOE) An annual Skills, Knowledge, and Responsibility Score (SKR score) is calculated for each teacher evaluated using data collected from evaluations using the TAP rubrics.

The principal summative evaluations were performed during the spring of 2013, and SKR scores were shared with the researcher by the principal based on the informed consent of each participating teacher.

#### **Phase 5: Value Added Measures Scores Data Collection**

The Value Added Measures Scores were based on the state of Louisiana's Compass (Appendix I) accountability matrix that calculates the score from high-stakes norm referenced and criterion referenced tests. Each individual

teacher receives a percentile score from 1- 100, and this score has a corresponding label as follows: 1%-10%= Ineffective, 11%-49% = Effective Emerging , 50%-89%= Effective Proficient, 90%-100%= Highly Proficient. Once an individual teacher's value-added percentile score has been tabulated, that score is then converted to the 1.00 to 4.00 Compass scaled score and it becomes the student growth measure that constitutes 50 percent of the teacher's final Compass rating.

The Value Added Measures were based on test results from the spring of 2013, and were released to the district in the summer of 2013. They were then shared with the researcher by the district testing coordinator based on the informed consent of each participating teacher on August 1<sup>st</sup>, 2013.

### **Data Analysis**

The unit of observation is the teacher. Teachers reported self-efficacy scores based on data collected in Phase 1. Demographic data was collected during Phase 2. Each teacher received a mean student perception score based on surveys administered to students in the spring of 2013 in Phase 3. Teachers received principal report scores based on observations conducted in the spring of 2013 and collected by the researcher during Phase 4. Teachers Value Added Measure (VAM) scores as measures of student achievement based on testing in spring of 2013 were collected in Phase 5. The researcher was interested understanding the relationship between these four variables.

The researcher used multiple regression to determine whether or not there was any association between these variables. In particular, the researcher

wanted to test the hypothesis as to whether the Value Added Scores of different teachers are statistically correlated with high and low student perceptions, high and low self-efficacy, and high and low principal evaluation.

The researcher wanted to fit a linear model to examine the distribution of the multi-dimensional data, and investigate the distribution. The approach was to fit a linear regression model to the data and examine the size of the coefficients.

The following objectives were developed by the researcher to accomplish the purpose of this study:

1. To describe teachers working in an elementary, middle, and high public school in the southeastern region of the United States on the following selected characteristics:
  - a. Years experience
  - b. Grades taught
  - c. Subjects taught
  - d. Gender
  - e. Race
  - f. Age

This objective is descriptive in nature, and therefore descriptive statistics were used to analyze the resulting data. Means, standard deviations and frequencies were used for analysis of demographic information.

2. To determine if there is a statistically significant relationship between the principal's/supervisor's observation score of a teacher and the

achievement of that teachers' students as measured by value-added standardized test scores.

To meet this objective, the teachers' observation score was correlated to his/her VAM score to determine if a relationship exists between the two variables.

3. To determine if there is a statistically significant relationship between the self-efficacy score of a teacher and the achievement of that teachers' students as measured by value-added standardized test scores.

To meet this objective, the teachers' self-efficacy score was correlated to his/her VAM score to determine if a relationship exists between the two variables.

4. To determine if there is a statistically significant relationship between the students' perceptions of engagement effectiveness of a teacher and the achievement of that teacher's students as measured by value-added standardized scores.

To meet this objective, the teachers' student perceptions score was correlated to his/her VAM score to determine if a relationship exists between the two variables.

For objectives 2, 3, and 4, Pearson's product moment correlation was used to determine if a relationship exists between the independent and dependent variables. Data analysis consisted of Pearson's Product Moment

Correlations. Since the primary variable of interest to the researcher was student achievement as measured by VAM scores, that score was forced into the multiple regression model first. The remaining variables were entered in a forward selection procedure because of the investigative nature of the study. The probability of  $F$  was set at .05 and the probability of  $F$  to be removed from the model was set at .10. The data was analyzed for normality, linearity, and homoscedasticity. Also, collinearity diagnostics and multiple regression diagnostics were analyzed to identify influential outliers.

To test the stated hypotheses, the preceding research study was conducted at a selected K-12 public school in the southeastern region of the United States. The unit of measure is the teacher. The study was conducted in the Spring of 2013 and analyzed during the summer and fall of that same year.

This correlational study was expected to yield powerful results due to the potential anticipated relationship between the independent variables (self-efficacy, observations, and student perceptions) and the directional correlations of student achievement as measured by value added test scores. If this result was observed, then it could be stated that students' perceptions of teacher effectiveness and measures of teachers' self-efficacy are greater predictors of student achievement than principal's observations. This result would have allowed the researcher to make predictive interpretations from the results of the study. Because this research is one of a limited number of correlational studies of its kind (only two were identified by this researcher through extensive review of literature, and none in this area of the United States), it is believed to add

knowledge to the limited body of research literature by documenting the statistical significance of relationship between multiple measures of teacher effectiveness and student achievement.

## **CHAPTER 4: PRESENTATION OF RESULTS**

### **Overview**

The purpose of this study was to compare three inputs of teacher evaluation – principals'/supervisors' observation scores, self-efficacy scores, and student perceptions of teacher effectiveness- to determine which is most strongly correlated to the academic outcomes of students as measured by value-added test scores. Three separate analyses of Pearson's Moment Correlation were conducted, and then a multiple regression analysis to determine the strongest relationship.

The study was carried out as a five phase study using quantitative measures. In phase one, the self-efficacy of each teacher was measured using the Teacher Sense of Efficacy Scale (Woolfolk-Hoy). In phase two, demographic data was collected from each respondent. Phase three consisted of the systematic administration of surveys designed to measure students' perceptions of teacher effectiveness. These surveys were aggregated into mean scores since the unit of analysis is the teacher. During phase four, the measures of principal observations were collected from the school administration. The fifth and final phase was the collection of VAM and SLT scores from the district testing coordinator.

This chapter presents the results of the analysis of the quantitative data from the case study school. The results are presented in the following order: 1.) Summary of response rates and total sample 2.) Results of analyses used to

address each of the research hypotheses posed in the first and third chapters, including descriptives and correlational data, and 3.) Additional results.

### **Summary of Response Rates and Total Sample**

68 teachers in the case study school agreed to participate in the study, for a total of 100%. All 68 completed the Teachers Sense of Efficacy Scale. Of these 68, 4 were removed from the sample for because they serve as master teachers or instructional coaches within the school and do not serve classes of students, so perceptions of their effectiveness could not be measured by student survey methods. High response rates can be credited to the dedicated cooperation of the school administration and the timeline of events. Self-efficacy surveys were administered during staff development day when the entire faculty was gathered together. Demographic data was gathered via electronic webform with follow-up reminders. Student perceptions measurements were gathered during a one week period in a coordinated effort involving the entire school. Principal observation scores and VAM/SLT scores were collected by reporting from the principal and the district testing coordinator, respectively.

While complete data sets of input variables were collected from 64 teachers, not all of these data sets were used in the analysis. 14 teachers were removed from the analysis because they taught elective subjects that are not measured by VAM scores or SLT scores in traditional ways (foreign language, music, physical education, fine arts, self-contained or inclusion special education, etc.). Upon review of the students' perceptions instrumentation, the thirteen Kindergarten, 1<sup>st</sup>, and 2<sup>nd</sup> grade teachers were also removed from analysis. The

researcher determined that there were items that could be unclear in their interpretation, and this could invalidate the results for this group. For example, the item “My teacher gives us homework” received several sad face responses. The directions instructed the students to mark a sad face if they disagree with the statement, but many students in follow-up discussions explained that they marked sad faces because homework makes them sad; not because they aren’t assigned any. There were several items that were miscoded for similar reasons (“our work is too hard for us”, “we do the same thing in class everyday”), so the K-2 data sets were omitted from analysis. The total sample data sets that were used in the analysis included 37 teachers.

Descriptive statistics relative to the sample are shown in Table 4.1. They are presented by teacher for those participating in the study that were included in the final analysis. Frequencies and percentages are given for all demographic variables (gender, ethnicity, age, years experience, grades taught, and subjects taught), and means and standard deviations are reported for age and years experience.

In Table 4.1, the demographic characteristics of the respondents in this study are also reported. Teachers were asked to report via webform a survey regarding various personal and professional factors. Females made up the majority of the respondent sample. Of the 37 teachers that were included in the data analysis, only 10.9 reported their gender as male. The racial makeup of the participants did not vary greatly, and 40.5% of the sample was non-white. The mean age of participants was 37.2 years with a standard deviation of 9.6, and the

mean years of professional teaching experience was 13.7 with a standard deviation of 9.5. The National Center for Educational Statistics reports that the average number of years of teaching experience of regular full-time public school teachers in the United States is 13.3.

Table 4.1. Demographics of Respondents used in data analysis

Demographic	n	Category	Frequency	Percent	Mean	SD
Gender	n=37	Male	4	10.9		
		Female	33	89.1		
Ethnicity	n=37	Black	13	35.1		
		White	22	59.5		
		Hispanic	1	2.7		
		Asian	1	2.7		
Age	n=37	22-29	6	16.2	37.2	9.6
		30-39	14	37.8		
		40-49	12	32.4		
		50-59	4	10.8		
		60+	1	2.7		
Years of Professional Teaching Experience	n=37	1-3	6	16.2	13.7	9.5
		4-5	4	10.8		
		6-10	7	18.9		
		11-20	12	32.4		
		21-30	7	18.9		
		30 or more	1	2.7		
Grade(s) Taught	n=37	3-5	14	37.8		
		6-8	10	27.0		
		9-12	13	35.1		
Subjects Taught	n=37	Math	5	13.5		
		English/LA	5	13.5		
		Science	4	10.8		
		Social Studies	2	5.4		
		More than one reported	21	56.8		

## **Results Addressing Research Hypotheses**

In the following section, results of correlational analyses of the independent and dependent variables are reported. Additionally, results addressing research questions are presented separately. The research hypotheses are posited at the beginning of the respective subsection, and are followed by explanation of the statistical analyses and discussion of results.

### **Research Hypothesis 1:**

Teachers working in an elementary, middle, and high (K-12) public school in the southeastern region of the United States with higher measures of self-efficacy will have higher value added test scores (as measured by VAM or SLT).

The researcher ran a bivariate correlation to calculate Pearson's  $r$  to determine whether or not a significant relationship existed between a teacher's total measure of Self Efficacy and his/her VAM scores. The value of this correlation for teachers whose student achievement is measured by VAM is  $r = .295$ , which is not significant at the predetermined .05 level.

The researcher then ran a bivariate correlation to calculate Pearson's  $r$  to determine whether or not a significant relationship existed between a teacher's total measure of self efficacy and his/her SLT scores. The value of this correlation is  $r = .205$ , which is not significant at the predetermined .05 level.

The results of these two correlational analyses show that there is no statistical relationship between a teacher's measure of self-efficacy and the student achievement of that teacher's students as measured by VAM scores or SLT scores. Based on these results, no predictive value on student achievement

as measured by the current VAM system can be assigned to the teacher's measure of self efficacy. Research hypothesis 1 is rejected based on these results.

### **Research Hypothesis 2:**

Teachers working in an elementary, middle, and high public school in the southeastern region of the United States with higher principal observation scores will have higher value added test scores.

The researcher ran a bivariate correlation to calculate Pearson's  $r$  to determine whether or not a significant relationship existed between the mean principals'/supervisors' observation scores (as measured by the Skills, Knowledge, and Responsibility Score) and his or her VAM scores. The value of this correlation is  $r = .567$ , which is highly significant at the .01 level.

The researcher then ran a bivariate correlation to calculate Pearson's  $r$  to determine whether or not a significant relationship existed between a teacher's mean principals'/supervisors' observation scores (as measured by the Skills, Knowledge, and Responsibility Score) and his or her SLT scores. The value of this correlation is  $r = .798$ , which is highly significant at the .01 level.

The results of these two correlational analyses show that there is a high statistically significant relationship between a teacher's mean principals'/supervisors' observation scores (as measured by the Skills, Knowledge, and Responsibility Score) and his or her student achievement as measured by VAM scores or SLT scores. Based on these results, principals'/supervisors observation measures of teachers may be interpreted as

strong predictors of student achievement as measured by VAM scores or SLT scores. Research hypothesis 2 is supported based on these results.

### **Research Hypothesis 3:**

Teachers working in an elementary, middle, and high public school in the southeastern region of the United States with higher scores in the students' perceptions of teacher effectiveness will have higher value added test scores. (as measured by VAM or SLT).

The researcher ran a bivariate correlation to calculate Pearson's  $r$  to determine whether or not a significant relationship existed between each individual teachers' mean score of effectiveness as measured by student perceptions surveys and his/her VAM scores. The value of this correlation for teachers whose student achievement is measured by VAM is  $r = .017$ , which is not significant at the predetermined .05 level.

The researcher then ran a bivariate correlation to calculate Pearson's  $r$  to determine whether or not a significant relationship existed each individual teacher's mean score of effectiveness as measured by student perceptions surveys and his/her SLT scores. The value of this correlation is  $r = .304$ , which is not significant at the predetermined .05 level.

The results of these two correlational analyses show that there is no statistical relationship between each individual teachers' mean score of effectiveness as measured by student perceptions surveys and the student achievement of that teacher's students as measured by VAM scores or SLT scores. Based on these results, no predictive value on student achievement as

measured by the current VAM system can be assigned to the students' perceptions of teacher effectiveness as measured by current instrumentation.

Research hypothesis 1 is rejected based on these results.

#### **Research Hypothesis 4:**

The students' perceptions of teacher effectiveness and the measures of self-efficacy will be more closely correlated to the value added test scores of the teacher than the principal observation.

A multiple regression analysis was deemed unnecessary based on results of statistical test run for RQ1, RQ2, and RQ3. Of the three independent variables (teachers' self-efficacy scores, principals/supervisors observation scores, and student perceptions of teacher effectiveness) that were measured, only one showed any significant relationship to the dependent variable of student achievement as measured by VAM scores or SLT scores.

These results indicate that the strongest relationship within and among these variables is the highly significant relationship between the principals/supervisors observations of teachers and the Value Added Measures of that teacher.

Research hypothesis 4 is rejected based on these results.

#### **Additional Results:**

The researcher ran correlational item analysis between each item on the student perceptions surveys for each group. For the 3-5 grade group and the 6-8 grade group, no significant relationships were found between any individual item and the VAM scores or SLT scores of the teacher. For the 9-12 group, there

were significant correlations at the .05 level between teachers' scores on items 4, 5, 6, and 9 and the teachers' SLT scores. Table 4.2 reports those results.

Table 4.2 Item by item analysis, grades 9-12 and SLT Scores

Item Number	Item	Pearson's <i>r</i>
item 4	We discuss and summarize each lesson we have studied.	.664*
item 5	My teacher tells us how we can use what we have already learned to learn new things.	.695*
item 6	My teacher maintains discipline in our classroom.	.730*
item 9	My teacher knows a lot about this subject.	.609*
item 19	My teacher is available to help me during class time and other times during the school day.	.136
item 16	My teacher likes it when we ask questions.	.186
item 18	My teacher encourages us to look at problems in new ways and find new ways to solve problems.	.216
item 17	We work in different groups depending upon the activity in which we are involved.	.367
item 11	My teacher makes materials and worksheets for us to use.	.386
item 7	My teacher returns tests and assignments quickly.	.400
item 1	My teacher makes our work interesting.	.406
item 14	My teacher tells the class about library/media materials that will help us learn about the subject we are studying, when appropriate.	.406
item 8	My teacher gives me feedback about my performance.	.443
item 10	My homework helps me to learn the subject being taught.	.457
item 20	My teacher looks at our work as we are DOIng it to see if we understand the lesson.	.458
item 13	The films or videos we watch help us learn about the subject we are studying.	.481
item 15	My teacher is well organized.	.489
item 12	My teacher uses a variety of classroom activities and resources.	.520
item 3	My teacher gives assignments related to the subject we are studying.	.581
item 2	My teacher asks questions to see if we understand what has been taught.	.596

## **CHAPTER 5: CONCLUSIONS**

### **Summary**

Teacher performance data were collected via a variety of reports: self-efficacy surveys, student performance surveys, and principal observation protocols all collected in the spring of 2013. Student achievement data were gathered in the form of Value-Added scores on state mandated tests. In all, 64 teachers participated in the study.

The study identified the relationship between the teacher performance evaluation measures (self, student, and principal) and student achievement in that teacher's classroom measured in terms of Value-Added Models. The researcher wanted to determine which rating would best predict the achievement of the teacher's students. The detailed findings which resulted from the hypothesis testing were presented in the previous chapter. The paramount conclusion of the study was that the principal's ratings of teachers are the best predictors of student achievement on standardized tests designed to measure the state mandated curriculum and reported using a Value-Added Model. Student ratings of teachers and measures of self-efficacy showed no statistical significance.

### **Implications, Recommendations, and Limitations**

The purpose of this study was to compare three inputs of teacher evaluation – principals'/supervisors' observation scores, self-efficacy scores, and student perceptions of teacher effectiveness- to determine which is most strongly correlated to the academic outcomes of students as measured by value-added

test scores. Such measures are needed to inform a total quality teacher evaluation system and to influence professional development decisions. Successive analyses of bivariate correlation coefficients were used to identify which measures were significantly related to student outcomes. The following section illustrates the conclusions, implications, and recommendations based on the findings discussed in Chapter Four. Discussion will also include how these findings relate to previous research.

### **Conclusion 1**

It was concluded that teachers working in an elementary, middle, and high (K-12) public school in the southeastern region of the United States with higher measures of self-efficacy did not have higher value added test scores (as measured by VAM or SLT) than other teachers. No predictive value on student achievement as measured by the current VAM system can be assigned to the teacher's measure of self efficacy. The results of this study differ from other studies which reported strongly significant relationships between measures of teachers' self-efficacy and student achievement. These results provide limited support for the hypothesized relationship between teachers' self-efficacy and student outcomes.

This information is of value to administrators of K-12 public schools in the southeastern region of the United States in determining which factors to include in their professional development process. With the current Value Added Measurement system in place, measures of self-efficacy are not related to student achievement using current instruments. As important as knowing what

measures to attend to in decision making is knowing what measures to not attend to; according to the results of this study, measures of self-efficacy can be disregarded as predictors of student achievement as they are currently recorded by the Compass System.

The review of related literature found information from three studies that reported that a teacher's measures of self-efficacy were very highly correlated to student achievement (Ashton and Webb, 1986; Ross, 1998; Isore, 2009). The results of this study are inconsistent with those findings. However, it is important to note that the measures of student achievement in other large-scale studies like this one---notably the MET project--- were not exclusively Value Added Models based on state standardized tests (Kane, 2012). Rather, the MET project measured student achievement in terms of state tests and supplemental standardized tests including the Stanford 9 Open-Ended Reading Assessment, the Balanced Assessment of Mathematics (BAM), and the ACT QualityCare series for Algebra I, English 9, and Biology. All of these assessments are norm-referenced. A possible explanation for the result could be the integrity of the Louisiana Value-Added Model in comparison to other models used in previous research.

A recommendation for further research is to investigate the same measure of self-efficacy and record results from various Value-Added Models that measure student achievement. While findings from this study indicate that a significant relationship does not exist, other studies using different barometers of

achievement have reported to the contrary. Another study could examine the differences in VAM types and self-efficacy results.

## **Conclusion 2**

It was concluded that teachers working in an elementary, middle, and high public school in the southeastern region of the United States with higher principal observation scores will have higher value added test scores. Based on these results, principals'/supervisors observation measures of teachers may be interpreted as strong predictors of student achievement as measured by VAM scores or SLT scores. The results of this study are consistent with other studies which reported strongly significant relationships between principal evaluation and student achievement. They are also consistent with current practices in teacher evaluation and professional development.

This information is of value to administrators and decision makers in K-12 public schools in the southeastern region of the United States in determining whether or not to continue with current practices of principal observations. The data indicates that results of principals' observations scores are indeed predictive of student achievement as measured by Value Added Measures. These issues are salient in regard to principal feedback, particularly when resources, rewards, and promotions resulting in merit pay increases for teachers are considered.

The implication of this finding is that current methods of principal evaluation can be useful in determining teacher quality and predicting student outcomes. An extensive review of the literature indicates that teachers largely view principal observation protocols as unproductive and lacking in value. The

findings in this study are interesting in that the teachers' perception is generally not supported by data.

An implication for further study could be to conduct further investigations into the methods and models of principal/supervisory observations and their relationships to student outcomes. Several observation protocols exist, and the determination and assessment of those with the highest predictive value is worthy of continued investigation.

### **Conclusion 3**

It was concluded that teachers working in an elementary, middle, and high (K-12) public school in the southeastern region of the United States with higher measures of effectiveness as reported through student perceptions surveys did not have higher value added test scores (as measured by VAM or SLT) than other teachers. No predictive value on student achievement as measured by the current VAM system can be assigned to the students' perceptions of teacher effectiveness. The results of this study differ from other studies which reported strongly significant relationships between measures of student feedback of effectiveness and student achievement. These results provide limited support for the hypothesized relationship between student feedback to teachers regarding effectiveness and student outcomes.

This information is of value to administrators of K-12 public schools in the southeastern region of the United States in determining which factors to include in their professional development process. With the current Value Added Measurement system in place, measures of effectiveness as reported by

students are not related to student achievement using current instruments. As stated in conclusion one, it is just as important to know what measures to attend to in decision making as to know what measures to ignore; according to the results of this study, student perceptions of teacher effectiveness can be disregarded as predictors of student achievement as they are currently recorded by the Compass System.

The review of related literature reported results from several researchers that student perceptions of teachers' effectiveness are strongly related to student outcomes (Omotani, 1996; Peck, 1998; Kane & Staiger, 2013). The results of this study are inconsistent with those findings. However, it is important to note that the measures of student achievement in other large-scale studies like this one--- notably the MET project--- were not exclusively Value Added Models based on state standardized tests (Kane, 2012). Rather, the MET project measured student achievement in terms of state tests and supplemental standardized tests including the Stanford 9 Open-Ended Reading Assessment, the Balanced Assessment of Mathematics (BAM), and the ACT QualityCare series for Algebra I, English 9, and Biology. All of these assessments are norm-referenced. A possible explanation for the result could be the integrity of the Louisiana Value-Added Model in comparison to other models used in previous research.

A recommendation for further research is to investigate the relationships between student perceptions of teacher effectiveness using other instruments and record results from various Value-Added Models that measure student achievement. While findings from this study indicate that a significant relationship

does not exist, other studies using different barometers of achievement have reported to the contrary. Another study could examine the differences in VAM types and student perceptions of effectiveness results.

### **Additional Results**

An item-by-item analysis demonstrated that four items on the 9-12 instrument were found to be significantly related to student achievement as measured by SLT scores. The correlations were noticeable at the .05 level, and the findings suggest that further attention in future studies should be given to these descriptors:

4) We discuss and summarize each lesson we have studied. ( $r = .664 > .05$ )

5) My teacher tells us how we can use what we have already learned to learn new things. ( $r = .695 > .05$ )

6) My teacher maintains discipline in our classrooms. ( $r = .730 > .05$ )

9) My teacher knows a lot about this subject. ( $r = .609 > .05$ )

Data indicates that teachers that demonstrate these behaviors have students with higher achievement. The implications of this conclusion may influence future lesson planning, sequencing, knowledge acquisition, and classroom management.

### **Limitations**

A number of limitations were imposed by the design of this study. They include:

1. The use of one school, while providing a comprehensive sample, resulted in some small sample sizes for each individual group (3-5, 6-8, 9-12).

Small (n) cell size may have affected the significance of differences between means of some groups. Larger (n) sizes may have changed some of the outcomes.

2. All teacher evaluation and student achievement data were collected during the 2012-13 school year. This fact prevented the analysis of findings in a longitudinal study beyond that time frame.
3. There are many variables not involved in the present study that likely affected both supervisor/principal ratings of teacher performance as well as student learning. A very likely one is the prior student performance of teachers influencing the principal's rating.
4. Only teachers in subject areas and grade levels that are measured by VAM scores or SLT scores were included in the study and subsequent data analysis. Caution should certainly be observed in generalizing these findings to other subject areas and grade levels.

Much research remains to be done to fully understand the predictors of teacher evaluation based on various measures. Researchers may want to improve and replicate this study by recognizing and modifying based on the above noted limitations. Future researchers may also want to improve upon the criterion used in this study, particularly those in the form of Value Added Scores/Student Learning Targets.

In conclusion, educational reformation and transition have always been recurring themes. During the past four decades, many strides have been made to improve the quality of K-12 education. Among these strides are increased

attention to the quality of teacher evaluation systems and continuous professional improvement. Many industries, businesses, and government organizations not employ multi-rater feedback as elements of evaluation systems. In schools, there is a climate of accountability which delineates student achievement as the most important yardstick to grade teacher performance. If the best criterion of effective teaching is indeed student performance, then it stands to reason that the students of the most effective teachers will learn more (Cashin, 1995). While this statement is seemingly obvious, teacher performance and evaluation remain contentious issues among educational stakeholders. The ways in which teacher performance actually relates to student performance has rarely been studied, but it should be as it can do much to corroborate assessment instruments to institute accountability. This study sought to provide some insight into the value of multiple rater feedback systems, particularly the contribution that students have to offer. Based on the literature review, it was expected that principal ratings, student ratings, and teacher self-ratings would all show positive relationships with student performance as measured by Value-Added scores. It was surprising to discover that the principal's ratings would be the only significant factor to predict student achievement of individual teachers. At the very least, the researcher felt that the student perceptions of teacher effectiveness would also show some degree of significance in predicting student outcomes based solely on the day-to-day exposure, but that was not the case. As observed in the current study, the strongest predictor of teacher effectiveness is the historical standard of principal observation. However, no individual

measure can ever encapsulate the total picture of a teacher's impact. Multiple rater systems provide instructors and school leaders with a better understanding of how teaching contributes to students' success. It will be critical that future studies be conducted to determine the most effective models for teacher evaluation and informed professional development.

## BIBLIOGRAPHY

- Adams, M. L. (2006). Quest for Tenure: Job Security and Academic Freedom. *Catholic University Law Review*, 56, 67. Retrieved from SSRN: <http://ssrn.com/abstract=1133641>.
- Aldeman, C., & Sector, E. (2011). *A Measured Approach to Improving Teacher Preparation*: Education Sector. Retrieved from <http://www.educationsector.org/publications/measured-approach-improving-teacher-preparation>.
- Anderson, J. D. (1988). *The education of Blacks in the South, 1860-1935*: JSTOR. Retrieved from [http://www.jstor.org/stable/10.5149/9780807898888\\_anderson](http://www.jstor.org/stable/10.5149/9780807898888_anderson).
- Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York, NY: Longman.
- Avalos, B., & Assael, J. (2006). Moving from resistance to agreement: The case of the Chilean teacher performance evaluation. *International Journal of Educational Research*, 45(4), 254-266. DOI:10.1016/j.ijer.2007.02.004.
- Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J., & Pincus, J. A. (1972). How Effective is Schooling?
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental psychology*, 25(5), 729. DOI:10.1037/0012-1649.25.5.729.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational psychologist*, 28(2), 117-148. DOI:10.1207/s15326985ep2802\_3.
- Bandura, A. (1994). *Self-efficacy*: Wiley Online Library. DOI: 10.1002/9780470479216.corpsy0836.
- Bass, B. M., & Stogdill, R. M. (1990). *Handbook of leadership. Theory, Research & Managerial Applications*. New York, NY: Free Press.
- Berry, K. S., & Herrington, C. D. (2011). States and Their Struggles With NCLB: Does the Obama Blueprint Get It Right? *Peabody Journal of Education*, 86(3), 272-290. DOI:10.1080/0161956X.2011.578982.
- Blake, R. R., & Mouton, J. S. (1968). *The Managerial Grid; Key Orientations for Achieving Production through People*. Miami, FL: Gulf Publishing.

- Braun, H. I. (2005). Value-added modeling: What does due diligence require. *Value-added models in education: Theory and applications*, 19-38.
- Buddin, R., McCaffrey, D. F., Kirby, S. N., & Xia, N. (2007). Merit pay for Florida teachers.
- Cashin, W. E. (1995). NOTE 10p.; Update of" Student Ratings of Teaching: A Summary of the Research. IDEA Paper No. 20,"see ED 302 567. *Center for Faculty Evaluation and Development*, 302, 567.
- Chan, K. (2001). The difficulties and dilemma of constructing a model for teacher evaluation in higher education. *ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT*, 13(1).
- Cremin, L. A. (1970). *American education: The colonial experience, 1607-1783* (Vol. 2): Harper & Row New York.
- Cruickshank, D. R., & Haefele, D. (2001). Good teachers, plural. *Educational Leadership*, 58(5), 26-30.
- Daley, G., & Kim, L. (2010). A teacher evaluation system that works. *National Institute for Excellence in Teaching*. Santa Monica CA.
- Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership*, 58(5), 12-15.
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.
- Danielson, C., & McGreal, T. (2000). Teacher evaluation to enhance professional learning. *Princeton, NJ: Educational Testing Service*.
- Danielson, C., & McGreal, T. L. (2005). *Teacher evaluation to enhance professional practice*: Ascd.
- Darling-Hammond, L. (2010). *The flat world and education*: Teachers College Press.
- Darling-Hammond, L., & Ducommun, C. (2010). Performance counts: Assessment systems that support high-quality learning. *Washington, Dc: council of chief State School officers*.
- Darling-Hammond, L., & Sykes, G. (2003). Wanted, A national teacher supply policy for education: The right way to meet the" highly qualified teacher" challenge. *education policy analysis archives*, 11, 33.

- Dolton, P., & Marcenaro-Gutierrez, O. D. (2011). If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic policy*, 26(65), 5-55. DOI:10.1111/j.1468-0327.2010.00257.x.
- Domas, S. J., & Tiedeman, D. V. (1950). Teacher competence: An annotated bibliography. *The Journal of Experimental Education*, 19(2), 101-218.
- Donaldson, M. L. (2009). So long, Lake Wobegon? Using teacher evaluation to raise teacher quality. *Center for American Progress*.
- Emmer, E. T., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement*, 51(3), 755-765. DOI:10.1177/0013164491513027.
- Force, C. T. (1986). A nation prepared: Teachers for the 21st century. *The report*.
- Garman, N. B. (1986). Reflection, the Heart of Clinical Supervision: A Modern Rationale for Professional Practice. *Journal of Curriculum and Supervision*, 2(1), 1-24.
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of educational psychology*, 76(4), 569. DOI:10.1037/0022-0663.76.4.569.
- Gilbert, T. F. (2007). *Human competence: Engineering worthy performance*: Wiley. com.
- Glanz, J. (2005). Action research as instructional supervision: Suggestions for principals. *NASSP Bulletin*, 89(643), 17-27. DOI:10.1177/019263650508964303.
- Goe, L. (2008). Key issue: Using value-added models to identify and support highly effective teachers. *Washington, DC: National Comprehensive Center for Teacher Quality*. Retrieved February, 28, 2010.
- Goe, L., Biggers, K., & Croft, A. (2012). Linking Teacher Evaluation to Professional Development: Focusing on Improving Teaching and Learning. Research & Policy Brief. *National Comprehensive Center for Teacher Quality*.
- Goldhammer, R. (1969). Clinical Supervision. New York: Holt, Rinehart and Winston. *Inc.(1969)*.
- Goldhammer, R., Anderson, R. H., & Krajewski, R. J. (1969). *Clinical supervision: Special methods for the supervision of teachers*: Holt, Rinehart and Winston New York.

- Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal*, 31(3), 627-643. DOI:10.3102/00028312031003627.
- Haefele, D. L. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education*, 7(1), 21-31. DOI:10.1007/BF00972346.
- Hazi, H. M., & Rucinski, D. A. (2005). Refocusing on the Ritual of Teacher Evaluation: Implications for the Future of Supervision. *Reflections of the Past, Visions for the Future*, 154.
- Holland, P. (2005). The case for expanding standards for teacher evaluation to include an instructional supervision perspective. *Journal of Personnel Evaluation in Education*, 18(1), 67-77. DOI:10.1007/s11092-006-9009-0.
- Huberman, M. (1985). What knowledge is of most worth to teachers? A knowledge-use perspective. *Teaching and Teacher Education*, 1(3), 251-262. DOI:10.1016/0742-051X(85)90008-3.
- Ingvarson, L., Kleinhenz, E., & Wilkinson, J. (2007). Research on performance pay for teachers.
- Ingvarson, L., & Rowe, K. (2008). Conceptualising and evaluating teacher quality: Substantive and methodological issues. *Australian Journal of Education*, 52(1), 5-35. DOI:10.1177/000494410805200102.
- Isoré, M. (2009). Current Practices in OECD Countries and a Literature Review. *OECD Education Working Paper*, 23(Directorate for Education). DOI:10.1787/223283631428.
- Jacob, A., Vidyarthi, E., & Carroll, K. (2012). The Irreplaceables: Understanding the Real Retention Crisis in America's Urban Schools. *TNTP*.
- Jacob, B. A., & Lefgren, L. (2005). Principals as agents: Subjective performance measurement in education: National Bureau of Economic Research. DOI:10.3386/w11463.
- Joyce, J. M., & Peck, J. (2005). *Creating a Community and Continuum of Teacher Education*. Paper presented at the Society for Information Technology & Teacher Education International Conference.
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. (2013). Have We Identified Effective Teachers? DOI:10.1596/1813-9450-6590.

- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project. *Bill & Melinda Gates Foundation*. DOI:10.1596/1813-9450-6590.
- Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation: National Bureau of Economic Research. DOI:10.3386/w14607.
- 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. MET Project. *Bill & Melinda Gates Foundation*.
- Kantor, H. (1991). Education, social reform, and the state: ESEA and federal education policy in the 1960s. *American Journal of Education*, 47-83. DOI:10.1086/444004.
- Kennedy, M. (2010). *Teacher assessment and the quest for teacher quality: A handbook*. Wiley. com.
- Kennedy, M. M. (2005). *Inside teaching: How classroom life undermines reform*. Harvard University Press.
- King, L. (2003). Teacher perceptions of the Danielson/McGreal model of differentiated evaluation and its impact on professional growth. *Unpublished doctoral dissertation, Seton Hall University, South Orange, NJ*.
- Kleinhenz, E., & Ingvarson, L. (2004). Teacher accountability in Australia: Current policies and practices and their relation to the improvement of teaching and learning. *Research papers in education*, 19(1). DOI:10.1080/0267152032000176963.
- Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, 76, 1-34.
- Korman, A. K. (1966). "Consideration," "Initiating Structure," and Organizational Criteria-a Review. *Personnel Psychology*, 19(4), 349-361. DOI:10.1111/j.1744-6570.1966.tb00310.x.
- Kupermintz, H. (2002). Value-added assessment of teachers. *School reform proposals: The research evidence*, 217-234.

- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational evaluation and policy analysis*, 25(3), 287-298. DOI:10.3102/01623737025003287.
- Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The future of children*, 17(1), 87-109. DOI:10.1353/foc.2007.0007.
- Liston, D., Whitcomb, J., & Borko, H. (2006). Too little or too much teacher preparation and the first years of teaching. *Journal of Teacher Education*, 57(4), 351-358. DOI:10.1177/0022487106291976.
- Marshall, K. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan*, 86(10), 727-735.
- Marshall, M. (2012). *Discipline Without Stress Punishments or Rewards: How Teachers and Parents Promote Responsibility & Learning*: Piper Press.
- McCaffrey, D. F., Lockwood, J., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating Value-Added Models for Teacher Accountability*. Monograph: ERIC.
- McCaffrey, D. F., Lockwood, J., 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. DOI:10.3102/10769986029001067.
- McGaghie, W. C. (1991). Professional competence evaluation. *Educational Researcher*, 20(1), 3-9. DOI:10.3102/0013189X020001003.
- McGreal, T. (1988). Evaluation for enhancing instruction: Linking teacher evaluation and staff development. *Teacher evaluation: Six prescriptions for success*, 1-29.
- McGuinn, P. (2012). The State of Teacher Evaluation Reform: State Education Agency Capacity and the Implementation of New Teacher-Evaluation Systems. *Center for American Progress*.
- Nowack, K. M. (1992). Self-assessment and rater-assessment as a dimension of management development. *Human Resource Development Quarterly*, 3(2), 141-155. DOI:10.1002/hrdq.3920030206.
- Nowack, K. M. (2009). Leveraging multirater feedback to facilitate successful behavioral change. *Consulting Psychology Journal: Practice and Research*, 61(4), 280. DOI:10.1037/a0017381.

- Nutt, H. W. (1920). *The supervision of instruction*: Houghton Mifflin.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational evaluation and policy analysis*, 26(3), 237-257. DOI:10.3102/01623737026003237.
- OECD (2013), OECD publishing. Teachers for the 21st Century: Using Evaluation to Improve Teaching.
- Omotani, B. J., & Omotani, L. M. (1996). Why Don't We Ask the Clients? Students' Input Toward a Complete Picture of Teaching. *NASSP Bulletin*, 80(577), 112-115. DOI:10.1177/019263659608057719.
- Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123-141.
- Peck, D. (1998). Teaching culture: Beyond language. *Yale: New Haven Teachers Institute*.
- Peterson, K. (2006). Using multiple data sources in teacher evaluation systems. *Evaluating teaching: A guide to current thinking and best practice*, 212-232.
- Peterson, K. D., Kelly, P., & Caskey, M. (2002). Ethical considerations for teachers in the evaluation of other teachers. *Journal of Personnel Evaluation in Education*, 16(4), 317-324. DOI:10.1023/A:1021752209823.
- Ponticell, J. A., & Zepeda, S. J. (2004). Confronting well-learned lessons in supervision and evaluation. *NASSP Bulletin*, 88(639), 43-59. DOI:10.1177/019263650408863905.
- Porter-Magee, K. (2004). Teacher quality, controversy, and NCLB. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(1), 26-29. DOI:10.3200/TCHS.78.1.26-29.
- Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74(6), 625-637. DOI:10.1002/sce.3730740605.
- Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education/Revue canadienne de l'education*, 51-65. DOI:10.2307/1495395.
- Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. *Advances in research on teaching*, 7, 49-74.

- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3), 393-404. DOI:10.5465/AMR.1998.926617.
- Savage, T. V., & McCord, M. K. (1986). The Use of Student Evaluation in the Assessment of Teacher Competence.
- Schurr, N., & Tambe, M. (2008). Using multi-agent teams to improve the training of incident commanders *Defence Industry Applications of Autonomous Agents and Multi-Agent Systems* (pp. 151-166): Springer. DOI:10.1007/978-3-7643-8571-2\_9.
- Shepherd, G. J., & Trank, D. M. (1989). Individual Differences in Consistency of Evaluation: Student Perceptions of Teacher Effectiveness. *Journal of Research and Development in Education*, 22(3), 45-52.
- Soodak, L. C., & Podell, D. M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education*, 12(4), 401-411. DOI:10.1016/0742-051X(95)00047-N.
- Spears, H. (1953). *Improving the supervision of instruction*: Prentice-Hall.
- Stronge, J., Helm, V., & Tucker, P. (1995). Evaluation handbook for professional support personnel: Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.
- Stronge, J. H. (1995). Balancing individual and institutional goals in educational personnel evaluation: A conceptual framework. *Studies in Educational Evaluation*, 21(2), 131-151. DOI:10.1016/0191-491X(95)00010-R.
- Stronge, J. H., & Hindman, J. L. (2003). Hiring the Best Teachers. *Educational Leadership*, 60(8), 48-52.
- Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*: Eye on Education Larchmont.
- Sykes, G., & Winchell, S. (2010). Assessing teacher tenure. *Handbook on Teacher Assessment and Teacher Quality*. San Francisco: Jossey-Bass.
- Toch, T., & Rothman, R. (2008). Rush to Judgment: Teacher Evaluation in Public Education. Education Sector Reports. *Education Sector*.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783-805. DOI:10.1016/S0742-051X(01)00036-1.

- Tyack, D. (1967). *Turning Points in American Education History*. Waltham, Mass: Blaisdell.
- Urban, W. J., & Wagoner, J. L. (2004). *American education: A history*. McGraw Hill Boston.
- Vinovskis, M. A. (2009). *From A Nation at Risk to No Child Left Behind: National education goals and the creation of federal education policy*.
- Vollmer, M. L., & Creek, R. J. (1988). Measurement of Student Perceptions of Teaching Competencies.
- Walsh, K., & Snyder, E. (2004). Searching the Attic: How States Are Responding to the Nation's Goal of Placing a Highly Qualified Teacher in Every Classroom. NCTQ Reports. *National Council on Teacher Quality*.
- Watt, H. M., & Richardson, P. W. (2008). Motivations, perceptions, and aspirations concerning teaching as a career for different types of beginning teachers. *Learning and Instruction*, 18(5), 408-428. DOI:10.1016/j.learninstruc.2008.06.002.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect. *Brooklyn, NY: The New Teacher Project*.
- Wiedmer, T. L. (1995). Instructional improvement: The goal of clinical supervision. *The Teacher Educator*, 30(3), 6-18. DOI:10.1080/08878739509555083.
- Wiggins, G. (1998). *Educative Assessment. Designing Assessments To Inform and Improve Student Performance*: ERIC.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*: Jossey-Bass.
- Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360 feedback® for teacher evaluation. *Journal of Personnel Evaluation in Education*, 14(2), 179-192. DOI:10.1023/A:1008158904681.

## **APPENDIX A CONSENT FORM**

### **Louisiana State University- Baton Rouge Campus Study Title: Using Self-Assessment of Efficacy and Student Feedback as a Component of Teacher Performance Evaluation Consent Form**

The purpose of this study is to investigate the benefits and drawbacks to including Self-Assessment of Efficacy and Student Feedback to Teachers as a component of an overall teacher performance evaluation system. We need volunteers to take part in this study and would like you to consider participating. Participation involves completion of the Teachers' Sense of Efficacy Scale (Moran and Woolfork-Hoy) and having students complete the Student Feedback to Teachers Instruments (Manatt). You will also be asked to answer follow-up questions for clarification and participate in focus group interviews. Your participation is entirely voluntary and you will not be penalized in any way for not volunteering.

Any discomforts or risks that may result from participation are minimal. Your participation will allow you to learn more about the ways that researchers attempt to reveal and understand important issues in education. The data gathered on you will be kept confidential, and any identifying information you provide will be omitted. All data will be examined only by duly authorized representatives of the research team and you are assured that the information will not be used for any purpose other than the scientific goals of the study. If you choose to participate, you are free to stop at any time without penalty of any sort.

Any questions you may have regarding procedures or any other aspect of the study can be answered by contacting Leslie Blanchard (225-716-9001).

This study has been approved by the LSU Institutional Review Board. If you have any questions about participants' rights you can contact Robert Mathews at 225-578-8692.

I give permission to the Principal Investigator, Leslie Blanchard, to use instrument responses collected from me and my students and informal responses as data collected for the purposes of her research study.

I have been briefed by the researcher in detail about this project and understand what my participation involves. I agree to participate with the understanding that I may withdraw at any time. I agree with the terms above and have read and understand this consent form.

\_\_\_\_\_  
Participant Signature

\_\_\_\_\_  
Today's Date

\_\_\_\_\_  
Print Name

## APPENDIX B PERMISSION TO USE TSES



ANITA WOOLFOLK HOY, Ph.D.

PROFESSOR  
PSYCHOLOGICAL STUDIES IN EDUCATION

Dear

You have my permission to use the *Teachers' Sense of Efficacy Scale* in your research. A copy of both the long and short forms of the instrument as well as scoring instructions can be found at:

<http://www.coe.ohio-state.edu/ahoy/researchinstruments.htm>

Best wishes in your work,

A handwritten signature in cursive script that reads "Anita Woolfolk Hoy".

Anita Woolfolk Hoy, Ph.D.  
Professor

COLLEGE OF EDUCATION  
29 WEST WOODRUFF AVENUE  
COLUMBUS, OHIO 43210-1177

[WWW.COE.OHIO-STATE.EDU/AHOY](http://WWW.COE.OHIO-STATE.EDU/AHOY)

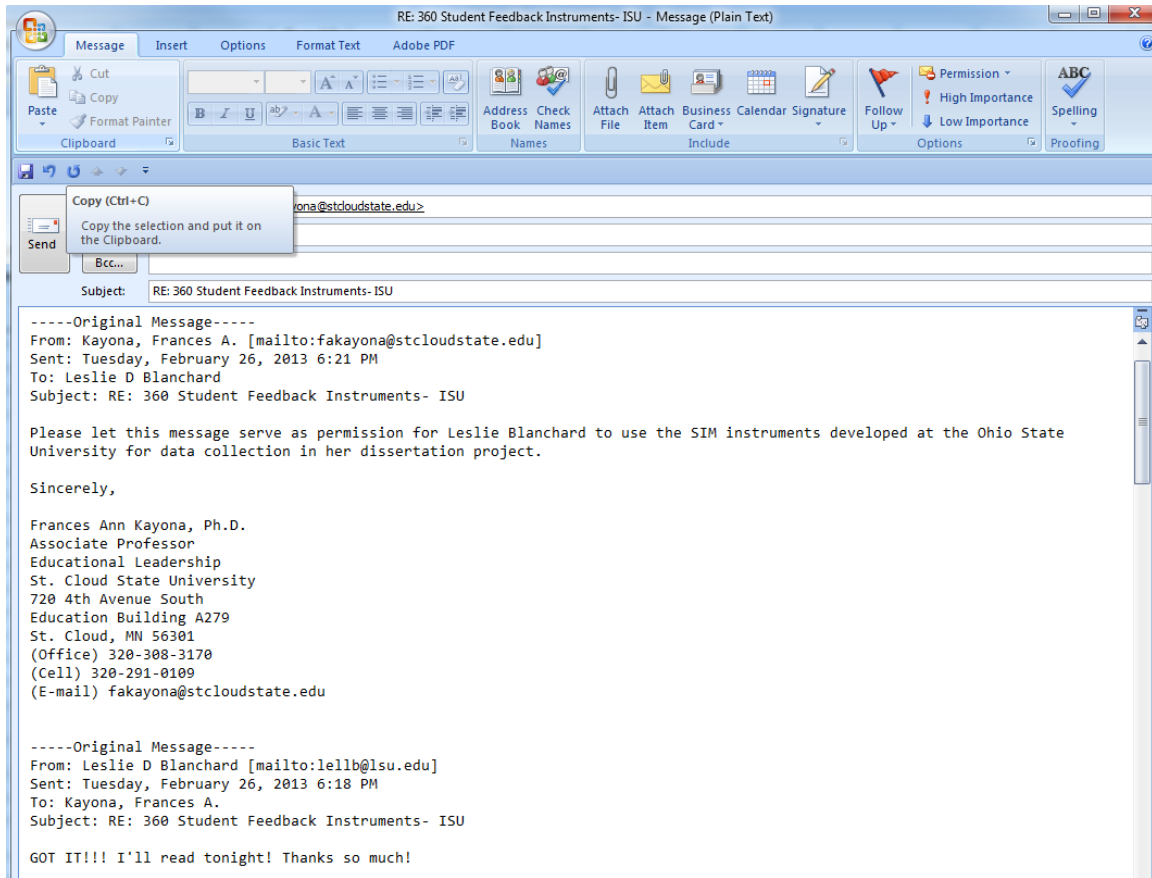
PHONE 614-292-3774  
FAX 614-292-7900  
HOY.17@OSU.EDU

## APPENDIX C TEACHERS SENSE OF EFFICACY SCALE

**Teachers' Sense of Efficacy Scale<sup>1</sup> (long form)**

<b>Teacher Beliefs</b>  <small>Directions: This questionnaire is designed to help us gain a better understanding of the kinds of things that create difficulties for teachers in their school activities. Please indicate your opinion about each of the statements below. Your answers are confidential.</small>	<b>How much can you do?</b>									
	Nothing	(1)	(2)	(3)	Very Little	(4)	Some influence	(5)	Quite A Bit	A Great Deal
1. How much can you do to get through to the most difficult students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
2. How much can you do to help your students think critically?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
3. How much can you do to control disruptive behavior in the classroom?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
4. How much can you do to motivate students who show low interest in school work?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
5. To what extent can you make your expectations clear about student behavior?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
6. How much can you do to get students to believe they can do well in school work?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
7. How well can you respond to difficult questions from your students ?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
8. How well can you establish routines to keep activities running smoothly?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
9. How much can you do to help your students value learning?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
10. How much can you gauge student comprehension of what you have taught?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
11. To what extent can you craft good questions for your students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
12. How much can you do to foster student creativity?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
13. How much can you do to get children to follow classroom rules?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
14. How much can you do to improve the understanding of a student who is failing?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
15. How much can you do to calm a student who is disruptive or noisy?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
16. How well can you establish a classroom management system with each group of students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
17. How much can you do to adjust your lessons to the proper level for individual students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
18. How much can you use a variety of assessment strategies?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
19. How well can you keep a few problem students from ruining an entire lesson?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
20. To what extent can you provide an alternative explanation or example when students are confused?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
21. How well can you respond to defiant students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
22. How much can you assist families in helping their children do well in school?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
23. How well can you implement alternative strategies in your classroom?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
24. How well can you provide appropriate challenges for very capable students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	

## APPENDIX D PERMISSION TO USE SIM INSTRUMENTS



## APPENDIX E SCRIPTS

MSA West Data Collection- April 30<sup>th</sup> and May 1<sup>st</sup>

Hello, and thank you once again for agreeing to participate in the data collection of student perceptions of teacher effectiveness. I really appreciate your cooperation!

The data collection should flow very smoothly and easily if everyone is prepared and we follow a few very simple instructions.

- 1.) Pick a “buddy teacher” to administer your surveys.

Teachers cannot survey their own students, so everyone will need to have a partner teacher to administer their instruments. Please identify a nearby colleague to pair with in survey administration, and a time when the two of you can “swap” classrooms for data collection. The instruments are all very short, and should not take more than 15 minutes to complete. Please make certain that all students understand that they are evaluating their teacher, NOT the person administering the survey.

- 2.) Make sure that you include the same class that was or will be used for your Compass Observation.

Each teacher is his/her own unit of study, and the correlations will be calculated back to his/her Value Added Scores or Student Learning Targets. The data will have more integrity if the students we collect surveys from are the same ones that were in the class from your observation. We will actually collect as much data as you like and report back on as many students as you requested, but for the sake of the study we will need to include that group at minimum.

- 3.) Please make sure that students code the Teacher ID number from the envelope provided onto the survey. That is the identifier associated with that teacher for all measures (self-efficacy, principal observation, and test scores), and will be confidential when results are reported.

\*\*\*THERE SHOULD BE NO TEACHER OR STUDENT NAMES ANYWHERE ON THE SHEETS!\*\*\*

- 4.) We want an accurate and honest depiction of student perceptions. For that reason, please try your best not to “coach” students to answer in a certain way. The following script can be used:

“Here is your chance to grade your teacher! Our school is participating in a study to evaluate the quality of our teaching staff. The survey that you are completing is an evaluation of Mr./Mrs./Ms. \_\_\_\_\_. Your name is not going to be recorded anywhere, and he/she will never know how you responded to this survey. Please be honest and answer each question as accurately as you can regarding your experience this school year in Mr./Mrs./Ms. \_\_\_\_\_’s class. The statements on the survey are designed to find out more about your class and teacher. This is not a test. Do not put your name on this paper. Please answer all the statements. Students are not allowed to ask any questions during the survey. If you have any questions, ask now.”

# APPENDIX F STUDENT PERCEPTIONS INSTRUMENTS- SIM

## XYZ ELEMENTARY SCHOOL Student Feedback to Teachers Lower Elementary School Questionnaire (K-2)

Please transfer the answers from each paper survey completed by the K-2 students to this form for scoring. Use the following scale:

0 = ☐ 1 = ☐ 2 = ☐

Make sure the teacher id is blackened.

1. My school day is interesting.
2. We do the same thing in class everyday.
3. I pay attention in class.
4. Our discussions are about the lesson being studied.
5. Our work is too hard for us.
6. My teacher gives us homework.
7. My teacher is usually prepared for class.
8. My teacher makes us follow the rules.
9. My teacher is fair with everybody.
10. My teacher cares if I waste time in class.
11. I work in this class even if the teacher is not watching.
12. I can get help from the teacher when I need it.
13. My teacher tells me I do good work.
14. My teacher tells me where I can find information to help me learn about the lesson.
15. My teacher is ready for class when it is time to begin.
16. I know what the teacher wants us to do.
17. My teacher is easy to understand.
18. My teacher has us learn hard lessons in small steps.
19. My teacher will explain new things in a way that is easy to learn.
20. My teacher tells us what new things we can learn in each lesson.

© 1994, Richard Marshall

TEACHER ID: \_\_\_\_\_

1. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
2. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
3. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
4. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
5. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
6. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
7. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
8. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
9. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
10. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
11. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
12. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
13. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
14. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
15. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
16. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
17. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
18. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
19. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4
20. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4

# XYZ ELEMENTARY SCHOOL

## Student Feedback to Teachers

### Upper Elementary School Questionnaire (3-5)

**NOTE TO STUDENTS:** Please remember that completing this form is voluntary. You may keep this form if you decide not to participate.

**Directions:** The statements below are designed to find out more about your class and teacher. This is not a test. Do not put your name on this paper. Please answer all the statements. Students are not allowed to ask any questions during the survey.

0 - Never 1 - Not often 2 - Sometimes 3 - Usually 4 - Almost always

1. My teacher makes our work interesting.
2. My school day is interesting.
3. We go back over each lesson when we finish it.
4. My teacher gives us work to do at home.
5. Our discussions are about the subject being studied.
6. My teacher gives our work back to us quickly.
7. My teacher makes me feel good when I do good work.
8. I can get help from my teacher.
9. I finish my work before class is over.
10. My teacher makes me follow the rules.
11. My teacher gives me new work to do without having to wait a long time for it.
12. My teacher explains the lesson clearly.
13. My teacher knows me well.
14. My teacher has work for me to do if I finish my assignment before class is over.
15. My teacher has us work at the right pace.
16. My teacher tells us what new things we can learn in each lesson.
17. My teacher will explain new things in a way that is easy to understand.
18. My teacher is available to help me during class time and other times during the school day.
19. My teacher uses a variety of classroom activities and resources.
20. My teacher is well prepared.

© 1994, Richard Manatt

TEACHER ID

0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4

1. 0 1 2 3 4
2. 0 1 2 3 4
3. 0 1 2 3 4
4. 0 1 2 3 4
5. 0 1 2 3 4
6. 0 1 2 3 4
7. 0 1 2 3 4
8. 0 1 2 3 4
9. 0 1 2 3 4
10. 0 1 2 3 4
11. 0 1 2 3 4
12. 0 1 2 3 4
13. 0 1 2 3 4
14. 0 1 2 3 4
15. 0 1 2 3 4
16. 0 1 2 3 4
17. 0 1 2 3 4
18. 0 1 2 3 4
19. 0 1 2 3 4
20. 0 1 2 3 4





## APPENDIX G STUDENT PERCEPTIONS INSTRUMENTS- LSU

### Kindergarten through Second Grade Student Perceptions Instrument

Please read each question to the students, and have them respond with a happy face if they agree with the statement, a sad face if they disagree with the statement, or a neutral face if they don't know which answer to choose.

1. My school day is interesting.
2. We do the same thing in class everyday.
3. I pay attention in class.
4. Our discussions are about the lesson being studied.
5. Our work is too hard for us.
6. My teacher gives us homework.
7. My teacher is usually prepared for class.
8. My teacher makes us follow the rules.
9. My teacher is fair with everybody.
10. My teacher cares if I waste time in class.
11. I work in this class even if the teacher is not watching.
12. I can get help from the teacher when I need it.
13. My teacher tells me I do good work.
14. My teacher tells me where I can find information to help me learn about the lesson.
15. My teacher is ready for class when it is time to begin.
16. I know what the teacher wants us to do.
17. My teacher is easy to understand.
18. My teacher has us learn hard lessons in small steps.
19. My teacher will explain new things in a way that is easy to learn.
20. My teacher tells us what new things we can learn in each lesson.

Teacher ID# 00-79

1. ☺☹☹
2. ☺☹☹
3. ☺☹☹
4. ☺☹☹
5. ☺☹☹
6. ☺☹☹
7. ☺☹☹
8. ☺☹☹
9. ☺☹☹
10. ☺☹☹
11. ☺☹☹
12. ☺☹☹
13. ☺☹☹
14. ☺☹☹
15. ☺☹☹
16. ☺☹☹
17. ☺☹☹
18. ☺☹☹
19. ☺☹☹
20. ☺☹☹



IBERVILLE MSA WEST MIDDLE SCHOOL  
Student Feedback to Teachers  
Middle School Questionnaire (6-8)

NOTE TO STUDENTS: Please remember that completing this form is voluntary. You may keep this form if you decide not to participate.

Directions: The statements below are designed to find out more about your class and teacher. This is not a test! Do not put your name on this paper. Please answer all the statements. Students are not allowed to ask any questions during the survey.

Teacher ID	
Room	Teacher
101	101
102	102
103	103
104	104
105	105
106	106
107	107
108	108
109	109
110	110
111	111
112	112
113	113
114	114
115	115
116	116
117	117
118	118
119	119
120	120

	Never	Not Often	Sometimes	Usually	Almost Always	
1. My teacher makes our work interesting.	00	00	00	00	00	00
2. My teacher is fair with all.	00	00	00	00	00	00
3. My teacher maintains discipline in our classroom.	00	00	00	00	00	00
4. My teacher is well prepared for our class.	00	00	00	00	00	00
5. My teacher gives assignments related to the subject we are studying.	00	00	00	00	00	00
6. We discuss and summarize each lesson studied.	00	00	00	00	00	00
7. Our discussions focus on the topic of the lesson.	00	00	00	00	00	00
8. My teacher likes it when we ask questions.	00	00	00	00	00	00
9. I have more time to do my work than I need.	00	00	00	00	00	00
10. My teacher starts lessons explaining what we are going to do and why we are going to do it.	00	00	00	00	00	00
11. My teacher asks a question in class to see if we understand what has been taught.	00	00	00	00	00	00
12. My teacher explains new ideas in a way that is easy to understand.	00	00	00	00	00	00
13. My teacher looks at our work, as we are doing it, to see if we understand the lesson.	00	00	00	00	00	00
14. My teacher is knowledgeable about his/her subject area.	00	00	00	00	00	00
15. My teacher has work for me to do if I finish an assignment before the class is over.	00	00	00	00	00	00
16. My teacher often makes materials and worksheets for us to use.	00	00	00	00	00	00
17. My teacher gives tests and quizzes.	00	00	00	00	00	00
18. My teacher returns tests and assignments quickly.	00	00	00	00	00	00
19. My teacher uses a variety of classroom activities and resources.	00	00	00	00	00	00
20. My teacher gives enough time to do our work.	00	00	00	00	00	00

81

4



## APPENDIX H TAP OBSERVATION RUBRIC

### *Teaching Skills, Knowledge, and Responsibilities Performance Standards*

The TAP *Teaching Skills, Knowledge, and Responsibilities Performance Standards* are the backbone of TAP's Instructionally Focused Accountability element. To measure teaching skills, knowledge, and responsibilities, one must define the skills and determine how they are demonstrated at different levels of performance. These standards were developed based on education psychology and cognitive science research focusing on learning and instruction, as well as an extensive review of publications from national and state teacher standards organizations.

The research for the *Teaching Skills, Knowledge, and Responsibilities Performance Standards* includes the following:

- » Milanowski, Odden & Youngs (1998) argue that the challenge of creating an effective teacher accountability system is to improve the quality of teacher instruction, and thereby raise student achievement. To do this, Odden and Clune (1998) instruct states and school districts to identify the knowledge and skills that a teacher needs to teach successfully, and then create standards and rubrics to measure teaching performance.
- » TAP reviewed instructional guidelines and standards developed by numerous national and state teacher standards organizations and from this information developed its own set of standards for teacher accountability. The work reviewed included guidelines and standards developed by:
  - The Interstate New Teacher Assessment and Support Consortium (INTASC)
  - The National Board for Professional Teacher Standards
  - Massachusetts's Principles for Effective Teaching
  - California's Standards for the Teaching Profession
  - Connecticut's Beginning Educator Support Program
  - The New Teacher Center's Developmental Continuum of Teacher Abilities
- » The criteria for the TAP teaching standards came from both experimental design studies and correlation studies that used valid and reliable achievement tests in classrooms (see Schacter & Thum, 2004).
- » The work of Danielson (1996) served as a valuable resource for defining the teaching competencies at each level of teacher performance.
- » Rubrics were designed based on the work of Rowley (1999) and various teacher accountability systems, including:
  - Rochester (New York) Career in Teaching Program
  - Douglas County (Colorado) Teacher's Performance Pay Plan
  - Vaughn Next Century Charter School (Los Angeles) Performance Pay Plan
  - Rolla (Missouri) School District Professional Based Teacher Evaluation

The TAP *Instruction, Designing and Planning Instruction, and The Learning Environment* rubrics are on the following pages. These rubrics and their 26 indicators are only intended for use by administrators, master teachers, and mentor teachers who have successfully completed their initial TAP evaluator certification and annual recertification. A more thorough explanation of the TAP Rubrics is found in the TAP Leadership Team Handbook. Following the rubrics is the evaluation report form used for teacher evaluations and self-evaluations.

## Teaching Skills, Knowledge, and Responsibilities Performance Standards Overview

INSTRUCTION	THE LEARNING ENVIRONMENT
<ol style="list-style-type: none"> <li>Standards and Objectives*</li> <li>Motivating Students*</li> <li>Presenting Instructional Content*</li> <li>Lesson Structure and Pacing*</li> <li>Activities and Materials*</li> <li>Questioning*</li> <li>Academic Feedback*</li> <li>Grouping Students*</li> <li>Teacher Content Knowledge*</li> <li>Teacher Knowledge of Students*</li> <li>Thinking*</li> <li>Problem Solving*</li> </ol>	<ol style="list-style-type: none"> <li>Expectations*</li> <li>Managing Student Behavior*</li> <li>Environment*</li> <li>Respectful Culture*</li> </ol>
DESIGNING AND PLANNING INSTRUCTION	RESPONSIBILITIES
<ol style="list-style-type: none"> <li>Instructional Plans</li> <li>Student Work</li> <li>Assessment</li> </ol>	<ol style="list-style-type: none"> <li>Staff Development**</li> <li>Instructional Supervision**</li> <li>Mentoring**</li> <li>Community Involvement**</li> <li>School Responsibilities**</li> <li>Growing and Developing Professionally</li> <li>Reflecting on Teaching</li> </ol>

\* Indicates criteria that are evaluated during classroom observations.

\*\* Indicates criteria that are only applied to master and mentor teachers.

Instruction			
	Exemplary (5)*	Proficient (3)*	Unsatisfactory (1)*
<b>Standards and Objectives</b>	<ul style="list-style-type: none"> <li>All learning objectives and state content standards are explicitly communicated.</li> <li>Sub-objectives are aligned and logically sequenced to the lesson's major objective.</li> <li>Learning objectives are: (a) consistently connected to what students have previously learned, (b) know from life experiences, and (c) integrated with other disciplines.</li> <li>Expectations for student performance are clear, demanding, and high.</li> <li>State standards are displayed and referenced throughout the lesson.</li> <li>There is evidence that most students demonstrate mastery of the objective.</li> </ul>	<ul style="list-style-type: none"> <li>Most learning objectives and state content standards are communicated.</li> <li>Sub-objectives are mostly aligned to the lesson's major objective.</li> <li>Learning objectives are connected to what students have previously learned.</li> <li>Expectations for student performance are clear.</li> <li>State standards are displayed.</li> <li>There is evidence that most students demonstrate mastery of the objective.</li> </ul>	<ul style="list-style-type: none"> <li>Few learning objectives and state content standards are communicated.</li> <li>Sub-objectives are inconsistently aligned to the lesson's major objective.</li> <li>Learning objectives are rarely connected to what students have previously learned.</li> <li>Expectations for student performance are vague.</li> <li>State standards are displayed.</li> <li>There is evidence that few students demonstrate mastery of the objective.</li> </ul>
<b>Motivating Students</b>	<ul style="list-style-type: none"> <li>The teacher consistently organizes the content so that it is personally meaningful and relevant to students.</li> <li>The teacher consistently develops learning experiences where inquiry, curiosity, and exploration are valued.</li> <li>The teacher regularly reinforces and rewards effort.</li> </ul>	<ul style="list-style-type: none"> <li>The teacher sometimes organizes the content so that it is personally meaningful and relevant to students.</li> <li>The teacher sometimes develops learning experiences where inquiry, curiosity, and exploration are valued.</li> <li>The teacher sometimes reinforces and rewards effort.</li> </ul>	<ul style="list-style-type: none"> <li>The teacher rarely organizes the content so that it is personally meaningful and relevant to students.</li> <li>The teacher rarely develops learning experiences where inquiry, curiosity, and exploration are valued.</li> <li>The teacher rarely reinforces and rewards effort.</li> </ul>
<b>Presenting Instructional Content</b>	<p>Presentation of content always includes:</p> <ul style="list-style-type: none"> <li>visuals that establish the purpose of the lesson, preview the organization of the lesson, and include internal summaries of the lesson;</li> <li>examples, illustrations, analogies, and labels for new concepts and ideas;</li> <li>modeling by the teacher to demonstrate his or her performance expectations;</li> <li>concise communication;</li> <li>logical sequencing and segmenting;</li> <li>all essential information and;</li> <li>no irrelevant, confusing, or nonessential information.</li> </ul>	<p>Presentation of content most of the time includes:</p> <ul style="list-style-type: none"> <li>visuals that establish the purpose of the lesson, preview the organization of the lesson, and include internal summaries of the lesson;</li> <li>examples, illustrations, analogies, and labels for new concepts and ideas;</li> <li>modeling by the teacher to demonstrate his or her performance expectations;</li> <li>concise communication;</li> <li>logical sequencing and segmenting;</li> <li>all essential information and;</li> <li>no irrelevant, confusing, or nonessential information.</li> </ul>	<p>Presentation of content rarely includes:</p> <ul style="list-style-type: none"> <li>visuals that establish the purpose of the lesson, preview the organization of the lesson, and include internal summaries of the lesson;</li> <li>examples, illustrations, analogies, and labels for new concepts and ideas;</li> <li>modeling by the teacher to demonstrate his or her performance expectations;</li> <li>concise communication;</li> <li>logical sequencing and segmenting;</li> <li>all essential information and;</li> <li>no irrelevant, confusing, or nonessential information.</li> </ul>
<b>Lesson Structure and Pacing</b>	<ul style="list-style-type: none"> <li>All lessons start promptly.</li> <li>The lesson's structure is coherent, with a beginning, middle, end, and time for reflection.</li> <li>Pacing is brisk and provides many opportunities for individual students who progress at different learning rates.</li> <li>Routines for distributing materials are seamless.</li> <li>No instructional time is lost during transitions.</li> </ul>	<ul style="list-style-type: none"> <li>Most lessons start promptly.</li> <li>The lesson's structure is coherent, with a beginning, middle, and end.</li> <li>Pacing is appropriate and sometimes provides opportunities for students who progress at different learning rates.</li> <li>Routines for distributing materials are efficient.</li> <li>Little instructional time is lost during transitions.</li> </ul>	<ul style="list-style-type: none"> <li>Lessons are not started promptly.</li> <li>The lesson has a structure, but may be missing closure or introductory elements.</li> <li>Pacing is appropriate for less than half of the students and rarely provides opportunities for students who progress at different learning rates.</li> <li>Routines for distributing materials are inefficient.</li> <li>Considerable time is lost during transitions.</li> </ul>

\* Performance definitions are provided at levels 5, 3, and 1. Raters can score performance at levels 2 or 4 based on their professional judgment.

## Instruction Continued

	Exemplary (5)	Proficient (3)	Unsatisfactory (1)
<b>Activities and Materials</b>	<p>Activities and materials include all of the following:</p> <ul style="list-style-type: none"> <li>• support the lesson objectives;</li> <li>• are challenging;</li> <li>• sustain students' attention;</li> <li>• elicit a variety of thinking;</li> <li>• provide time for reflection;</li> <li>• are relevant to students' lives;</li> <li>• provide opportunities for student-to-student interaction;</li> <li>• induce student curiosity and suspense;</li> <li>• provide students with choices;</li> <li>• incorporate multimedia and technology and;</li> <li>• incorporate resources beyond the school curriculum texts (e.g., teacher-made materials, manipulatives, resources from museums, cultural centers, etc.).</li> </ul> <p>In addition, sometimes activities are game-like, involve simulations, require creating products, and demand self-direction and self-monitoring.</p>	<p>Activities and materials include most of the following:</p> <ul style="list-style-type: none"> <li>• support the lesson objectives;</li> <li>• are challenging;</li> <li>• sustain students' attention;</li> <li>• elicit a variety of thinking;</li> <li>• provide time for reflection;</li> <li>• are relevant to students' lives;</li> <li>• provide opportunities for student-to-student interaction;</li> <li>• induce student curiosity and suspense;</li> <li>• provide students with choices;</li> <li>• incorporate multimedia and technology and;</li> <li>• incorporate resources beyond the school curriculum texts (e.g., teacher-made materials, manipulatives, resources from museums, cultural centers, etc.).</li> </ul>	<p>Activities and materials include few of the following:</p> <ul style="list-style-type: none"> <li>• support the lesson objectives;</li> <li>• are challenging;</li> <li>• sustain students' attention;</li> <li>• elicit a variety of thinking;</li> <li>• provide time for reflection;</li> <li>• are relevant to students' lives;</li> <li>• provide opportunities for student-to-student interaction;</li> <li>• induce student curiosity and suspense;</li> <li>• provide students with choices;</li> <li>• incorporate multimedia and technology and;</li> <li>• incorporate resources beyond the school curriculum texts (e.g., teacher-made materials, manipulatives, resources from museums, etc.).</li> </ul>
<b>Questioning</b>	<p>Teacher questions are varied and high quality, providing a balanced mix of question types:</p> <ul style="list-style-type: none"> <li>◦ knowledge and comprehension;</li> <li>◦ application and analysis; and</li> <li>◦ creation and evaluation.</li> </ul> <ul style="list-style-type: none"> <li>• Questions are consistently purposeful and coherent.</li> <li>• A high frequency of questions is asked.</li> <li>• Questions are consistently sequenced with attention to the instructional goals.</li> <li>• Questions regularly require active responses (e.g., whole class signaling, choral responses, written and shared responses, or group and individual answers).</li> <li>• Wait time (3-5 seconds) is consistently provided.</li> <li>• The teacher calls on volunteers and nonvolunteers, and a balance of students based on ability and sex.</li> <li>• Students generate questions that lead to further inquiry and self-directed learning.</li> </ul>	<p>Teacher questions are varied and high quality, providing for some, but not all, question types:</p> <ul style="list-style-type: none"> <li>◦ knowledge and comprehension;</li> <li>◦ application and analysis; and</li> <li>◦ creation and evaluation.</li> </ul> <ul style="list-style-type: none"> <li>• Questions are usually purposeful and coherent.</li> <li>• A moderate frequency of questions asked.</li> <li>• Questions are sometimes sequenced with attention to the instructional goals.</li> <li>• Questions sometimes require active responses (e.g., whole class signaling, choral responses, or group and individual answers).</li> <li>• Wait time is sometimes provided.</li> <li>• The teacher calls on volunteers and nonvolunteers, and a balance of students based on ability and sex.</li> </ul>	<p>Teacher questions are inconsistent in quality and include few question types:</p> <ul style="list-style-type: none"> <li>◦ knowledge and comprehension;</li> <li>◦ application and analysis; and</li> <li>◦ creation and evaluation.</li> </ul> <ul style="list-style-type: none"> <li>• Questions are random and lack coherence.</li> <li>• A low frequency of questions is asked.</li> <li>• Questions are rarely sequenced with attention to the instructional goals.</li> <li>• Questions rarely require active responses (e.g., whole class signaling, choral responses, or group and individual answers).</li> <li>• Wait time is inconsistently provided.</li> <li>• The teacher mostly calls on volunteers and high ability students.</li> </ul>

National Institute for Excellence in Teaching. Do not duplicate without permission.

## Instruction Continued

	Exemplary (5)	Proficient (3)	Unsatisfactory (1)
<b>Academic Feedback</b>	<ul style="list-style-type: none"> <li>Oral and written feedback is consistently academically focused, frequent, and high quality.</li> <li>Feedback is frequently given during guided practice and homework review.</li> <li>The teacher circulates to prompt student thinking, assess each student's progress, and provide individual feedback.</li> <li>Feedback from students is regularly used to monitor and adjust instruction.</li> <li>Teacher engages students in giving specific and high-quality feedback to one another.</li> </ul>	<ul style="list-style-type: none"> <li>Oral and written feedback is mostly academically focused, frequent, and mostly high quality.</li> <li>Feedback is sometimes given during guided practice and homework review.</li> <li>The teacher circulates during instructional activities to support engagement and monitor student work.</li> <li>Feedback from students is sometimes used to monitor and adjust instruction.</li> </ul>	<ul style="list-style-type: none"> <li>The quality and timeliness of feedback is inconsistent.</li> <li>Feedback is rarely given during guided practice and homework review.</li> <li>The teacher circulates during instructional activities, but monitors mostly behavior.</li> <li>Feedback from students is rarely used to monitor or adjust instruction.</li> </ul>
<b>Grouping Students</b>	<ul style="list-style-type: none"> <li>The instructional grouping arrangements (either whole class, small groups, pairs, or individual; heterogeneous or homogeneous ability) consistently maximize student understanding and learning efficiency.</li> <li>All students in groups know their roles, responsibilities, and group work expectations.</li> <li>All students participating in groups are held accountable for group work and individual work.</li> <li>Instructional group composition is varied (e.g., race, gender, ability, and age) to best accomplish the goals of the lesson.</li> <li>Instructional groups facilitate opportunities for students to set goals, reflect on, and evaluate their learning.</li> </ul>	<ul style="list-style-type: none"> <li>The instructional grouping arrangements (either whole class, small groups, pairs, or individual; heterogeneous or homogeneous ability) adequately enhance student understanding and learning efficiency.</li> <li>Most students in groups know their roles, responsibilities, and group work expectations.</li> <li>Most students participating in groups are held accountable for group work and individual work.</li> <li>Instructional group composition is varied (e.g., race, gender, ability, and age) to, most of the time, accomplish the goals of the lesson.</li> </ul>	<ul style="list-style-type: none"> <li>The instructional grouping arrangements (either whole class, small groups, pairs, or individual; heterogeneous or homogeneous ability) inhibit student understanding and learning efficiency.</li> <li>Few students in groups know their roles, responsibilities, and group work expectations.</li> <li>Few students participating in groups are held accountable for group work and individual work.</li> <li>Instructional group composition remains unchanged, irrespective of the learning and instructional goals of a lesson.</li> </ul>
<b>Teacher Content Knowledge</b>	<ul style="list-style-type: none"> <li>Teacher displays extensive content knowledge of all the subjects she or he teaches.</li> <li>Teacher regularly implements a variety of subject-specific instructional strategies to enhance student content knowledge.</li> <li>The teacher regularly highlights key concepts and ideas and uses them as bases to connect other powerful ideas.</li> <li>Limited content is taught in sufficient depth to allow for the development of understanding.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher displays accurate content knowledge of all the subjects he or she teaches.</li> <li>Teacher sometimes implements subject-specific instructional strategies to enhance student content knowledge.</li> <li>The teacher sometimes highlights key concepts and ideas and uses them as bases to connect other powerful ideas.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher displays under-developed content knowledge in several subject areas.</li> <li>Teacher rarely implements subject-specific instructional strategies to enhance student content knowledge.</li> <li>Teacher does not understand key concepts and ideas in the discipline and therefore presents content in an unconnected way.</li> </ul>
<b>Teacher Knowledge of Students</b>	<ul style="list-style-type: none"> <li>Teacher practices display understanding of each student's anticipated learning difficulties.</li> <li>Teacher practices regularly incorporate student interests and cultural heritage.</li> <li>Teacher regularly provides differentiated instructional methods and content to ensure children have the opportunity to master what is being taught.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher practices display understanding of some students' anticipated learning difficulties.</li> <li>Teacher practices sometimes incorporate student interests and cultural heritage.</li> <li>Teacher sometimes provides differentiated instructional methods and content to ensure children have the opportunity to master what is being taught.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher practices demonstrate minimal knowledge of students' anticipated learning difficulties.</li> <li>Teacher practices rarely incorporate student interests or cultural heritage.</li> <li>Teacher practices demonstrate little differentiation of instructional methods or content.</li> </ul>

Instruction Continued			
	Exemplary (5)	Proficient (3)	Unsatisfactory (1)
<b>Thinking</b>	<p>Over the course of multiple observations, the teacher consistently and thoroughly teaches all four types of thinking:</p> <ul style="list-style-type: none"> <li>analytical thinking, where students analyze, compare and contrast, and evaluate and explain information;</li> <li>practical thinking, where students use, apply, and implement what they learn in real-life scenarios;</li> <li>creative thinking, where students create, design, imagine, and suppose and;</li> <li>research-based thinking, where students explore and review a variety of ideas, models, and solutions to problems.</li> </ul> <p>The teacher regularly provides opportunities where students:</p> <ul style="list-style-type: none"> <li>generate a variety of ideas and alternatives;</li> <li>analyze problems from multiple perspectives and viewpoints and;</li> <li>monitor their thinking to ensure that they understand what they are learning, are attending to critical information, and are aware of the learning strategies that they are using and why.</li> </ul>	<p>Over the course of multiple observations, the teacher consistently and thoroughly teaches two types of thinking:</p> <ul style="list-style-type: none"> <li>analytical thinking, where students analyze, compare and contrast, and evaluate and explain information;</li> <li>practical thinking, where students use, apply, and implement what they learn in real-life scenarios;</li> <li>creative thinking, where students create, design, imagine, and suppose and;</li> <li>research-based thinking, where students explore and review a variety of ideas, models, and solutions to problems.</li> </ul> <p>The teacher sometimes provides opportunities where students:</p> <ul style="list-style-type: none"> <li>generate a variety of ideas and alternatives and;</li> <li>analyze problems from multiple perspectives and viewpoints.</li> </ul>	<p>The teacher implements few learning experiences that thoroughly teach any type of thinking.</p> <p>The teacher provides few opportunities where students:</p> <ul style="list-style-type: none"> <li>generate a variety of ideas and alternatives and;</li> <li>analyze problems from multiple perspectives and viewpoints.</li> </ul> <p>NOTE: If the teacher regularly and thoroughly teaches one type of thinking, he or she shall receive a score of 2.</p>
<b>Problem Solving</b>	<p>Over the course of multiple observations the teacher implements activities that teach and reinforce 6 or more of the following problem-solving types.</p> <ul style="list-style-type: none"> <li>Abstraction</li> <li>Categorization</li> <li>Drawing Conclusions/Justifying Solutions</li> <li>Predicting Outcomes</li> <li>Observing and Experimenting</li> <li>Improving Solutions</li> <li>Identifying Relevant/Irrelevant Information</li> <li>Generating Ideas</li> <li>Creating and Designing</li> </ul>	<p>Over the course of multiple observations the teacher implements activities that teach and reinforce 4 or more of the following problem-solving types.</p> <ul style="list-style-type: none"> <li>Abstraction</li> <li>Categorization</li> <li>Drawing Conclusions/Justifying Solution</li> <li>Predicting Outcomes</li> <li>Observing and Experimenting</li> <li>Improving Solutions</li> <li>Identifying Relevant/Irrelevant Information</li> <li>Generating Ideas</li> <li>Creating and Designing</li> </ul>	<p>Over the course of multiple observations the teacher implements less than 2 activities that teach the following problem-solving types.</p> <ul style="list-style-type: none"> <li>Abstraction</li> <li>Categorization</li> <li>Drawing Conclusions/Justifying Solution</li> <li>Predicting Outcomes</li> <li>Observing and Experimenting</li> <li>Improving Solutions</li> <li>Identifying Relevant/Irrelevant Information</li> <li>Generating Ideas</li> <li>Creating and Designing</li> </ul>

## Designing and Planning Instruction

	Exemplary (5)	Proficient (3)	Unsatisfactory (1)
<b>Instructional Plans</b>	<p>Instructional plans include:</p> <ul style="list-style-type: none"> <li>measurable and explicit goals aligned to state content standards;</li> <li>activities, materials, and assessments that:               <ul style="list-style-type: none"> <li>are aligned to state standards.</li> <li>are sequenced from basic to complex.</li> <li>build on prior student knowledge, are relevant to students' lives, and integrate other disciplines.</li> <li>provide appropriate time for student work, student reflection, and lesson and unit closure;</li> </ul> </li> <li>evidence that plan is appropriate for the age, knowledge, and interests of all learners and;</li> <li>evidence that the plan provides regular opportunities to accommodate individual student needs.</li> </ul>	<p>Instructional plans include:</p> <ul style="list-style-type: none"> <li>goals aligned to state content standards;</li> <li>activities, materials, and assessments that:               <ul style="list-style-type: none"> <li>are aligned to state standards.</li> <li>are sequenced from basic to complex.</li> <li>build on prior student knowledge.</li> <li>provide appropriate time for student work, and lesson and unit closure;</li> </ul> </li> <li>evidence that plan is appropriate for the age, knowledge, and interests of most learners and;</li> <li>evidence that the plan provides some opportunities to accommodate individual student needs.</li> </ul>	<p>Instructional plans include:</p> <ul style="list-style-type: none"> <li>few goals aligned to state content standards;</li> <li>activities, materials, and assessments that:               <ul style="list-style-type: none"> <li>are rarely aligned to state standards.</li> <li>are rarely logically sequenced.</li> <li>rarely build on prior student knowledge</li> <li>inconsistently provide time for student work, and lesson and unit closure;</li> </ul> </li> <li>little evidence that the plan is appropriate for the age, knowledge, or interests of the learners and;</li> <li>little evidence that the plan provides some opportunities to accommodate individual student needs.</li> </ul>
<b>Student Work</b>	<p>Assignments require students to:</p> <ul style="list-style-type: none"> <li>organize, interpret, analyze, synthesize, and evaluate information rather than reproduce it;</li> <li>draw conclusions, make generalizations, and produce arguments that are supported through extended writing and;</li> <li>connect what they are learning to experiences, observations, feelings, or situations significant in their daily lives, both inside and outside of school.</li> </ul>	<p>Assignments require students to:</p> <ul style="list-style-type: none"> <li>interpret information rather than reproduce it;</li> <li>draw conclusions and support them through writing and;</li> <li>connect what they are learning to prior learning and some life experiences.</li> </ul>	<p>Assignments require students to:</p> <ul style="list-style-type: none"> <li>mostly reproduce information;</li> <li>rarely draw conclusions and support them through writing and;</li> <li>rarely connect what they are learning to prior learning or life experiences.</li> </ul>
<b>Assessment</b>	<p>Assessment Plans:</p> <ul style="list-style-type: none"> <li>are aligned with state content standards;</li> <li>have clear measurement criteria;</li> <li>measure student performance in more than three ways (e.g., in the form of a project, experiment, presentation, essay, short answer, or multiple choice test);</li> <li>require extended written tasks;</li> <li>are portfolio-based with clear illustrations of student progress toward state content standards and;</li> <li>include descriptions of how assessment results will be used to inform future instruction.</li> </ul>	<p>Assessment Plans:</p> <ul style="list-style-type: none"> <li>are aligned with state content standards;</li> <li>have measurement criteria;</li> <li>measure student performance in more than two ways (e.g., in the form of a project, experiment, presentation, essay, short answer, or multiple choice test);</li> <li>require written tasks and;</li> <li>include performance checks throughout the school year.</li> </ul>	<p>Assessment Plans:</p> <ul style="list-style-type: none"> <li>are rarely aligned with state content standards;</li> <li>have ambiguous measurement criteria;</li> <li>measure student performance in less than two ways (e.g., in the form of a project, experiment, presentation, essay, short answer, or multiple choice test) and;</li> <li>include performance checks, although the purpose of these checks is not clear.</li> </ul>

## The Learning Environment

	Exemplary (5)	Proficient (3)	Unsatisfactory (1)
<b>Expectations</b>	<ul style="list-style-type: none"> <li>Teacher sets high and demanding academic expectations for every student.</li> <li>Teacher encourages students to learn from mistakes.</li> <li>Teacher creates learning opportunities where all students can experience success.</li> <li>Students take initiative and follow through with their own work.</li> <li>Teacher optimizes instructional time, teaches more material, and demands better performance from every student.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher sets high and demanding academic expectations for every student.</li> <li>Teacher encourages students to learn from mistakes.</li> <li>Teacher creates learning opportunities where most students can experience success.</li> <li>Students complete their work according to teacher expectations.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher expectations are not sufficiently high for every student.</li> <li>Teacher creates an environment where mistakes and failure are not viewed as learning experiences.</li> <li>Students demonstrate little or no pride in the quality of their work.</li> </ul>
<b>Managing Student Behavior</b>	<ul style="list-style-type: none"> <li>Students are consistently well-behaved and on task.</li> <li>Teacher and students establish clear rules for learning and behavior.</li> <li>The teacher uses several techniques, such as social approval, contingent activities, and consequences to maintain appropriate student behavior.</li> <li>The teacher overlooks inconsequential behavior.</li> <li>The teacher deals with students who have caused disruptions rather than the entire class.</li> <li>The teacher attends to disruptions quickly and firmly.</li> </ul>	<ul style="list-style-type: none"> <li>Students are mostly well-behaved and on task, some minor learning disruptions may occur.</li> <li>Teacher establishes rules for learning and behavior.</li> <li>The teacher uses some techniques, such as social approval, contingent activities, and consequences to maintain appropriate student behavior.</li> <li>The teacher overlooks some inconsequential behavior, but other times addresses it, stopping the lesson.</li> <li>The teacher deals with students who have caused disruptions, yet sometimes he or she addresses the entire class.</li> </ul>	<ul style="list-style-type: none"> <li>Students are not well-behaved and are often off task.</li> <li>Teacher establishes few rules for learning and behavior.</li> <li>The teacher uses few techniques to maintain appropriate student behavior.</li> <li>The teacher cannot distinguish between inconsequential behavior and inappropriate behavior.</li> <li>Disruptions frequently interrupt instruction.</li> </ul>
<b>Environment</b>	<p>The classroom</p> <ul style="list-style-type: none"> <li>welcomes all members and guests.</li> <li>is organized and understandable to all students.</li> <li>supplies, equipment, and resources are easily and readily accessible.</li> <li>displays student work that frequently changes.</li> <li>is arranged to promote individual and group learning.</li> </ul>	<p>The classroom</p> <ul style="list-style-type: none"> <li>welcomes most members and guests.</li> <li>is organized and understandable to most students.</li> <li>supplies, equipment, and resources are accessible.</li> <li>displays student work.</li> <li>is arranged to promote individual and group learning.</li> </ul>	<p>The classroom</p> <ul style="list-style-type: none"> <li>is somewhat cold and uninviting.</li> <li>is not well organized and understandable to students.</li> <li>supplies, equipment, and resources are difficult to access.</li> <li>does not display student work.</li> <li>is not arranged to promote group learning.</li> </ul>
<b>Respectful Culture</b>	<ul style="list-style-type: none"> <li>Teacher-student interactions demonstrate caring and respect for one another.</li> <li>Students exhibit caring and respect for one another.</li> <li>Teacher seeks out and is receptive to the interests and opinions of all students.</li> <li>Positive relationships and interdependence characterize the classroom.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher-student interactions are generally friendly, but may reflect occasional inconsistencies, favoritism, or disregard for students' cultures.</li> <li>Students exhibit respect for the teacher and are generally polite to each other.</li> <li>Teacher is sometimes receptive to the interests and opinions of students.</li> </ul>	<ul style="list-style-type: none"> <li>Teacher-student interactions are sometimes authoritarian, negative, or inappropriate.</li> <li>Students exhibit disrespect for the teacher.</li> <li>Student interaction is characterized by conflict, sarcasm, or put-downs.</li> <li>Teacher is not receptive to interests and opinions of students.</li> </ul>

## Evaluator/Self-Evaluation Report

- ☐ Announced  
☐ Unannounced

Evaluator \_\_\_\_\_ Administrator Master Mentor

Teacher Evaluated \_\_\_\_\_

Date \_\_\_\_\_ Time \_\_\_\_\_ Subject \_\_\_\_\_

School Name \_\_\_\_\_ Cycle 1 2 3 4 5 6

Designing and Planning Instruction	Evaluator Scores	Self-Eval Scores	Reinforcement Objective
Instructional Plans (IP)			
Student Work (SW)			
Assessment (AS)			
<b>The Learning Environment</b>			
Expectations (ES)			
Managing Student Behavior (MSB)			
Environment (ENV)			
Respectful Culture (RC)			
<b>Instruction</b>			<b>Refinement Objective</b>
Standards and Objectives (S&O)			
Motivating Students (MOT)			
Presenting Instructional Content (PIC)			
Lesson Structure and Pacing (LS)			
Activities and Materials (ACT)			
Questioning (QU)			
Academic Feedback (FEED)			
Grouping Students (GRP)			
Teacher Content Knowledge (TCK)			
Teacher Knowledge of Students (TKS)			
Thinking (TH)			
Problem Solving (PS)			

Evaluator Signature \_\_\_\_\_ Date \_\_\_\_\_

Teacher Signature \_\_\_\_\_ Date \_\_\_\_\_

## APPENDIX I COMPASS ACCOUNTABILITY MATRIX



### Compass End-of-Year Process Guide

Compass is a tool intended to support educators' professional growth, so that students meet new, rigorous [academic standards](#). Administrators, teachers, and counselors have used Compass to set [student learning targets](#), [observe classroom practice](#), and provide feedback on areas of success and areas of improvement.

As the school year comes to a close, teachers and their evaluators should come together to assess student learning and to agree on areas of strength and areas for improvement. These year-end conversations should also inform teachers' preparation for the coming school year by identifying specific areas for their own development and tools that will support their instructional planning and execution. This document is intended to serve as a resource to educators and their evaluators as they prepare for this end-of-year process.

The [Classroom Support Toolbox](#) should also be a helpful resource in these conversations. The Toolbox has been expanded to include [assessment guides](#), [sample year-long plans](#), and [classroom videos](#). Educators will use these tools to establish long-term, unit, and lesson plans for next school year.

Administrators who use the Compass tool successfully will conduct end-of-year discussions that reflect on student success, assess classroom practice, and use the [Classroom Support Toolbox](#) and other resources to plan for the upcoming year.

#### Timing

Soon you will also be able to access the final evaluation rating you receive from your school's administration. You may also look it up in the [Compass Information System](#), but administrators who use the Compass tool well will also share your rating with you through discussion and explanation. Your district and school will set the timelines, but all final evaluation ratings are to be completed by July 29.

Upcoming Compass Items	Action	Dates
Student Roster Verification	<ul style="list-style-type: none"><li>Teachers and school leaders verify student rosters in the value-added system</li></ul>	Start: April 8 End: May 10
End-of-Year Conversations	<ul style="list-style-type: none"><li>Evaluators conduct end-of-year conversations with educators</li></ul> <p>*Value-added data will not be available until the week of June 24. District and school leaders should develop a plan for discussing value-added results with teachers and leaders.</p>	Scheduled by evaluators
Value-Added Data Available	<ul style="list-style-type: none"><li>Teachers and school leaders review results</li><li>Evaluators submit final student outcomes rating for value-added teachers between 20<sup>th</sup> and 80<sup>th</sup> percentiles</li></ul>	June 27 – July 22
Final Evaluation Data Due in Compass Information System	<ul style="list-style-type: none"><li>Districts are required to submit all remaining final evaluation data by this time</li></ul>	July 29

Revised 7-8-13

## FREQUENTLY ASKED QUESTIONS

*All questions denoted with \* were collected during the Compass session at the Teacher Leader Kickoff on Friday, April 19, 2013.*

### Final Evaluation Ratings

#### 1. Who receives a Compass final evaluation rating?

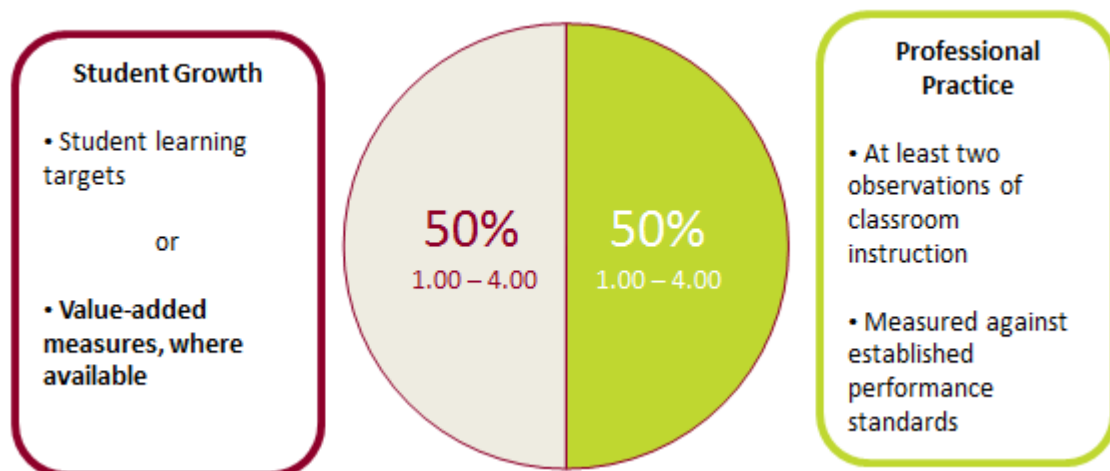
All teachers and administrators in Louisiana public schools will receive a final Compass rating for the 2012-13 school year. Teacher and administrator are defined as:

**Teacher:** Any person who provides direct instruction or direct instructional support to students to whom he/she has been assigned. This includes classroom teachers, librarians, and professional school counselors.

**Administrator:** Any person who serves in an academic leadership role at the school level and is employed in a professional capacity other than a teacher. This includes principals, assistant principals, and academic deans.

#### 2. How are the ratings calculated?

Teachers and school leaders receive an overall Compass score that is a combination of the student growth scores and professional practice scores.



**The student growth score** ranges between 1.00 and 4.00. This score is based on evidence from student learning targets and/or value-added data, according to the chart below.

## Educator Group

- School leaders
- Teachers who are not eligible to receive value-added data

## Student Growth Evidence used in final evaluation

Student learning targets

Value-added results

- Teachers who receive value-added results at the 20<sup>th</sup> percentile or below or at the 80<sup>th</sup> percentile or above

- Teachers who receive value-added results within the *Effective* range, between the 20<sup>th</sup> and 80<sup>th</sup> percentiles

Evaluator's assessment of both value-added data and student learning targets

here. Note: Guidance for evaluators on assessing student growth for teachers in the *Effective* range is available

here. Note: Further explanation of the *Effective* range is available

**The professional practice score** also ranges between 1.00 and 4.00. This rating is based on evidence collected during observations and site visits throughout the year. If the evaluator conducted more than two observations or site visits, she decides which of those are included in the final evaluation. For each observation or site visit included in the final evaluation, the scores for each component of the teacher or leader rubric are averaged to yield an overall observation/site visit score. Those scores are averaged at the end of the year to determine the final professional practice rating.

**The final Compass score** is calculated by averaging the professional practice score and the student growth score. This overall score between 1.00 and 4.00 corresponds with a final effectiveness rating, according to the chart below. Scores are automatically rounded to the nearest hundredths place in the Compass Information System.

Effectiveness Rating	Composite Score Range
<i>Ineffective</i>	$x < 1.50$
<i>Effective: Emerging</i>	$1.50 \leq x < 2.50$
<i>Effective: Proficient</i>	$2.50 \leq x < 3.50$
<i>Highly Effective</i>	$3.50 \leq x$

### 3. Who is responsible for calculating my final rating?

Your evaluator will enter ratings into the Compass Information System for your individual student learning targets and each component of your observations. The Compass Information System will automatically calculate your final Compass score according to the process described above, based on the data entered by your evaluator.



## **VITA**

Leslie Davis Blanchard was born in Thibodaux, Louisiana on to Alvin Henry Davis Jr. and Charlene Waguespack Davis. She graduated from Central Lafourche High School in 1991 and received a Bachelor of Science degree, majoring in secondary science education and minoring in English education, from Nicholls State University. While working as a teacher at White Castle High School, she won both school and district Teacher of the Year honors. Leslie returned to school to pursue her Master of Education in Administration and Supervision, graduating in 2001 from Southern University- Baton Rouge (SUBR). In 2002, Leslie was the recipient of the Presidential Award for Excellence in Math and Science Teaching which is awarded each year by the National Science Foundation (NSF) to one outstanding science and math teacher in each state. She received her award from then President George W. Bush.

From 2000-2009 Leslie served as the Instructional Technology Coordinator for the Iberville Parish School System. During her tenure there, she wrote and was awarded over \$12 million in both competitive and allocated grant funding at the state and federal level. From 2009 to the present, she has served as the Operations Manager for the Louisiana Math and Science Teachers' Institute (LaMSTI) at Louisiana State University (LSU). Leslie's goal of earning her doctoral degree from LSU was a major factor involved in her transition from the K-12 education sector to higher education. She began pursuit of the degree in the fall of 2011, and completed her coursework in the fall of 2013. Leslie and

her children Cassidy Anne Blanchard and Caroline Elizabeth Blanchard live in Brusly, Louisiana.