

Louisiana State University

## LSU Scholarly Repository

---

LSU Doctoral Dissertations

Graduate School

---

2016

# The Influence of Cultural and Contextual Factors on the Scoring Decisions of Rural Louisiana Principals during the Implementation of the COMPASS System for Teacher Evaluation

Keith Michael Courville

*Louisiana State University and Agricultural and Mechanical College*

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



Part of the [Education Commons](#)

---

### Recommended Citation

Courville, Keith Michael, "The Influence of Cultural and Contextual Factors on the Scoring Decisions of Rural Louisiana Principals during the Implementation of the COMPASS System for Teacher Evaluation" (2016). *LSU Doctoral Dissertations*. 2038.

[https://repository.lsu.edu/gradschool\\_dissertations/2038](https://repository.lsu.edu/gradschool_dissertations/2038)

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

THE INFLUENCE OF CULTURAL AND CONTEXTUAL FACTORS ON THE  
SCORING DECISIONS OF RURAL LOUISIANA PRINCIPALS DURING THE  
IMPLEMENTATION OF THE COMPASS SYSTEM FOR TEACHER EVALUATION

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 Department of Educational Leadership, Research, & Counseling

by  
Keith Michael Courville  
B.S. Louisiana College, 2006  
M.Ed., Louisiana State University, 2010  
Ed.S., Louisiana State University, 2012  
May 2016

## **ACKNOWLEDGEMENTS**

I would like to acknowledge the support of my wife, Dr. Rosemary Courville and my family, including my immediate family and grandparents: Jimmy, Catherine, Scott, Earl, and Patricia Courville, as well as Ferdinand Sr., Ferdinand Jr., and Merlene Verret. I would also like to especially thank Dr. Kenneth (Kenny) Fasching-Varner for his support and encouragement, as well as Dr. Roland Mitchell and Dr. Lori Martin for serving as my committee members, and Dr. Dana Bickmore for her work throughout the process.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
ABSTRACT .....	v
CHAPTER 1: HISTORY AND ISSUES .....	1
Introduction .....	1
Federal Interventions .....	2
Teacher Evaluation in Louisiana: An Overview .....	6
Response to RTTT .....	8
Statement of the Problem .....	9
Rationale for the Study .....	11
Conceptual Framework .....	11
Purpose Statement .....	13
Research Questions .....	13
Definition of Terms .....	14
Delimitations and Limitations .....	15
Position of the Researcher .....	16
Chapter Summary .....	16
CHAPTER 2: LITERATURE REVIEW .....	18
Introduction .....	18
Teacher Evaluation Systems .....	21
Teacher Evaluation in Louisiana .....	42
Implementation of Teacher Evaluation Systems .....	55
Conceptual Framework: Culture and Context .....	66
Summary .....	81
CHAPTER 3: METHODOLOGY .....	83
Introduction .....	83
Epistemological Foundation. ....	83
Research Questions .....	84
Setting .....	85
Participants .....	86
Data Collection .....	87
Data Analysis .....	88
Summary .....	91
CHAPTER 4: RESULTS .....	92
Participants .....	92
General Perspectives on the COMPASS Rubric .....	92
Training and Support During Implementation .....	96
Understanding of Scoring Expectations .....	98

External Factors and Their Influence .....	102
Scoring Teachers as Ineffective .....	106
Summary .....	109
CHAPTER 5: DISCUSSION .....	110
Summary .....	110
Conclusion .....	119
Suggestions for Future Research .....	119
Concluding Discussion .....	120
REFERENCES .....	122
VITAE .....	135

## **ABSTRACT**

As a result of the federal grant competition and policy initiative known as Race to the Top (RTTT), states began an examination and reform of the methods by which they evaluated teachers. Upon examination of the historical results of a multitude of evaluation systems, it was revealed that the majority of teachers score in the upper echelons of their performance reviews, in sharp contrast to the level of academic achievement or academic growth of students of US students.

After many reforms and the implementation of new teacher evaluation systems, such as the Compass system in Louisiana, the same positive skew that historically existed was once again found in the current teacher scores. In Louisiana, principals in almost half of the school districts did not rate a single teacher as ineffective. This phenomenon was most pronounced in rural school districts.

From the testimonies of principals in rural school districts, principals allowed other pressing concerns, namely the context of the rural environment and the culture of their school, to influence their scoring decisions when evaluating teachers.

## **CHAPTER 1: HISTORY AND ISSUES**

### **Introduction**

American policymakers continue to seek educational reforms, often through accountability policies, for what they view as unacceptable academic performance gaps (Linn, Baker, & Betebenner, 2002). For example, performance by American students on internationally benchmarked assessments, such as the Program for International Student Assessment (PISA), lags behind the performance of other developed nations (Fleischman, Hopstock, Pelczar, & Shelley, 2010). In Louisiana, students have consistently scored lower than their American peers on assessments of academic performance, such as the National Assessment of Educational Progress (NAEP) (Mullis, 1993; Szymanski, 2010). Although critics of American education reform denounce the existence of academic performance gaps as attempts to justify the undermining of traditional systems of public education, policymakers continue to develop and implement accountability policies in various areas (Bracey, 1996; Linn et al., 2002).

Proponents of accountability have shown an increasing interest in teacher evaluation. Traditionally, local school administrators determined teacher evaluation policies and practices. During the 1980s, some state policymakers recommended accountability policies in the field of teacher evaluation (McGreal, 1983; Wise, 1984). According to Danielson and McGreal (2000), researchers in the 1990s sought to improve teacher evaluation practices as a means to improve the academic outcomes of students. In a historical review of the policy field, Ellett and Teddlie (2003) also noted a significant paradigm shift during the 1990s, when both researchers and policymakers began to associate effective teaching with student learning. According to Ellett and Teddlie (2003), researchers developed evaluation instruments that incorporated certain standards

of professional practice associated with increased student performance, such as effective classroom management and the effective use of classroom time; these rubrics are known as standards-based rubrics. At the turn of the century, investigators began to examine standards of effective instructional practice and the use of standards-based rubrics (Darling-Hammond, 1999; Ladson-Billings & Darling-Hammond, 2000).

Although state policymakers developed and implemented teacher evaluation systems, the resulting scoring distributions were often positively skewed, with nearly all teachers receiving high scores (Danielson & McGreal, 2000). Danielson and McGreal (2000) explained this positive skew as possibly resulting from a lack of standards describing effective teaching or a cultural phenomenon aimed at protecting teachers from negative consequences. Ho and Kane (2013) described a need for adequately trained evaluators to reduce the positive skew. Researchers also found that teacher performance on standards-based rubrics correlated weakly or moderately with measures of student learning (Borman & Kimball, 2005; Milanowski, Kimball, & White, 2004). The research developments related to teacher evaluation provided a policy structure by which federal policymakers under the Obama administration could intervene in state accountability policy.

### **Federal Interventions**

Until the 1960s, education policymaking was the responsibility of individual states (Hubsch, 1992). Federal policymakers began to intervene in state education policy during President Lyndon Johnson's administration through the Elementary and Secondary Education Act of 1965 (ESEA). The ESEA contained provisions to provide incentives, often in the form of federal funding, for states' leaders to improve the level of academic achievement among impoverished students (Wargo, 1972). After its initial passage, federal policymakers continually



reauthorized the ESEA, making minor changes until the 1990s. One example of this reauthorization practice is the Education Consolidation and Improvement Act of 1981 (ECIA), which reorganized the federal programs section of the ESEA (Darling-Hammond & Marks, 1983).

In 1983, Federal policymakers intensified their role in education with the publication of *A Nation At Risk*. The report's authors presented the American education system as inadequate, citing how poorly American students performed compared to their international peers (Gardner, 1983). The report identified specific weaknesses in the coursework and assessments taken by American students; the report's authors pushed American policymakers to redesign coursework and increase high school graduation requirements (Gardner, 1983). The federal interventions spurred by *A Nation At Risk* set a precedent for future interventions in shaping the education policy of individual states (Guthrie & Springer, 2004).

In 1994, the Clinton administration passed the Improving America's Schools Act (IASA) (Johnson, 1997; Leonard, 1994). IASA provided funding to schools through President Clinton's reauthorization of the ESEA (Le Tendre, 1996). However, the provisions of IASA expanded the scope of federal funding and supported anti-drug campaigns, educational technology, and academic intervention programs to raise student achievement in underperforming populations; IASA also provided a mechanism for federal oversight of student assessment (Leonard, 1994), which established a framework for future federal accountability policies. Specifically, IASA required states to assess student performance and to evaluate schools (Goertz, 2005). However, IASA allowed state policymakers to define performance goals for schools and students, which resulted in varying student performance goals among states (Goertz, 2005).

Federal policymakers continued to evolve and to expand their accountability efforts through the precedent set by the IASA (McDonnell, 2005). In 2001, President George W. Bush signed No Child Left Behind (NCLB) into law as a reauthorization of the ESEA (Debray, McDermott, & Wohlstetter, 2005). Through this reauthorization, policymakers used the assessment guidelines of IASA to set school accountability goals that were uniform across the nation (Goertz, 2005). The provisions of NCLB held schools districts accountable for making improvements in student achievement on required annual assessments through an accountability policy known as annual yearly progress (AYP) (Linn et al., 2002). NCLB required AYP for all students, with the eventual goal of all students in all schools attaining proficiency by 2014 (Kim & Sunderman, 2005; Linn et al., 2002).

After NCLB under the Bush administration, President Barack Obama sought to further expand the federal role in education and abandoned the practice of reauthorizing the ESEA in favor of direct federal action (Manna & Ryan, 2011). In 2009, the Obama administration created Race to the Top (RTTT), a grant competition for state policymakers. A component of the American Recovery and Reinvestment Act (ARRA), RTTT was “designed to encourage and reward states that [were] creating the conditions for education innovation and reform” (USDOE, 2009, p. 2). The provisions of RTTT went beyond the focus on standards, assessment, and accountability found in the IASA and NCLB; RTTT promoted new areas of federal intervention, including efforts to address teacher quality and evaluation (Superfine, Gottlieb, & Smylie, 2012). The federal government allocated and distributed a total of 4.35 billion dollars to competing states based on how state policymakers proposed to meet certain criteria (USDOE, 2009). The four major areas of reform and their required policy initiatives are listed and briefly described below.

1. Standards and Assessments: evidence that “internationally benchmarked” (USDOE, 2009, p. 8) standards and assessments aligned with those standards were under development.
2. Data Systems to Support Instruction: evidence of “a statewide longitudinal data system” (USDOE, 2009, p. 8) that teachers and administrators would use to inform instructional practices.
3. Great Teachers and Leaders: evidence of “improving teachers and principal effectiveness based on performance” (USDOE, 2009, p. 9).
4. Turning Around the Lowest-Achieving Schools: evidence of a plan that described how the state would support local districts in implementing “one of...the four [prescribed] school intervention models” (USDOE, 2009, p. 10).

Relevant to the present study was the criteria related to the reform area of Great Teachers and Leaders, specifically “improving teacher and principal effectiveness based on performance” (USDOE, 2009, p. 9). This criterion contained several sub-criteria, and including the three relevant sub-criteria below:

- i. Establish clear approaches to measuring student growth...and measure it for each individual student;
- ii. Design and implement rigorous, transparent, and fair evaluation systems for teachers ... that (a) differentiate effectiveness using multiple rating categories that take into account data on student growth ... as a significant factor, and (b) are designed and developed with teacher and principal involvement;
- iii. Conduct annual evaluations of teachers and principals that include timely and constructive feedback; as part of such evaluations, provide teachers and

principals with data on student growth for their students, classes, and schools (USDOE, 2009, p. 9).

These sub-criteria are pertinent to this study as they apply to how principals evaluate teachers.

Federal policymakers included teacher evaluation in the RTTT selection criteria, which represented a developing consensus on the need to improve teacher evaluation systems by tying the performance ratings of teachers to student learning (Superfine et al., 2012). Before RTTT, local and state policymakers enacted laws establishing teacher evaluation systems as an effort to address public demand for accountability (Danielson & McGreal, 2000). However, prior to RTTT there was no mandate for teacher evaluation systems to “take into account data on student growth” (USDOE, 2009, p. 9), a stipulation which the above sub-criteria ii and iii included.

Discussed more fully in Chapter 2, many states, including Louisiana, had to revise their teacher evaluation policies and procedures to be eligible for RTTT funding (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). The following section explains how teacher evaluation policy developed in Louisiana.

### **Teacher Evaluation in Louisiana: An Overview**

Before RTTT, Louisiana policymakers developed and implemented two statewide teacher evaluation systems. The following passages present an overview of these systems and their associated instruments. Specific aspects of these systems and their instruments are detailed in Chapter 2.

Under the administration of Governor Buddy Roemer, legislators passed the Children First Act of 1988, which required principals to evaluate teachers and the results of which were used by the Louisiana Department of Education (LDOE) for the granting of certification, recertification, and tenure (Baldwin, 1995). In response to the Children First Act of 1988, the

LDOE policymakers implemented a statewide teacher evaluation system known as the Louisiana Teacher Evaluation Program (LATEP) (Johnson Jr., 1999). The LDOE designed LATEP to incorporate the system of rewards and consequences outlined by the legislature (Baldwin, 1995; Johnson Jr., 1999). For example, under LATEP, policymakers provided incentives such as additional compensation, or merit pay, for teachers with high scores on their evaluations (Baldwin, 1995). Under LATEP, principals used a classroom observation instrument known as the System for Teaching and Learning Assessment and Review (STAR) to provide a performance score for each teacher. Chad Ellett of Louisiana State University led the development and pilot testing of STAR. Ultimately, both LATEP and STAR faced criticism from teachers, unions, and politicians during their first year of implementation in the 1990-1991 school year (Johnson Jr., 1999). For example, teachers expressed frustrations about the lack of teacher participation in the development of the STAR instrument (Baldwin, 1995).

Baldwin (1995) chronicled the policy history in Louisiana during the administration of Governor Roemer including the implementation and subsequent cancelation of LATEP. The controversy over LATEP and STAR resulted in lawsuits against the Children First Act of 1998, legislative hearings, and protests by teachers (Baldwin, 1995). To address the controversy over LATEP and STAR, the Board of Elementary and Secondary Education (BESE) established a new instrument known as the Louisiana Components of Effective Teaching (LCET) (Baldwin, 1995). The LCET was developed by “a panel of educators consisting of teachers, principals, superintendents, and college faculty” (Baldwin, 1995, p. 118). Ultimately, BESE replaced LATEP, and its program of consequences and rewards for teachers, with the LCET (Baldwin, 1995). There is, however, no indication or evidence that BESE pilot-tested the LCET for validity or reliability before replacing STAR.

## **Response to RTTT**

Neither the LATEP system and its associated STAR instrument nor the LCET instrument met the criteria policymakers established for the RTTT competition. Neither STAR nor the LCET incorporated student performance into ratings of teacher performance. To be considered for RTTT funding, the Louisiana legislature enacted Act 54 of 2010, which codified the RTTT criteria pertaining to educator evaluation (Act No. 54, 2010). Act 54 required annual classroom observations of teachers and for those annual evaluations to include student test scores. Pursuant to this act, the Louisiana Department of Education (LDOE) created a new evaluation system, known as COMPASS. The COMPASS system contained two equally rated components: a quantitative component measuring student growth and a qualitative component generated from classroom observations. The quantitative component measured student academic growth through either the state's value-added measure (VAM), when appropriate standardized test score data exist, or teacher-developed goals for student learning, known as student learning targets (SLTs) (LDOE, 2013b). The LDOE only calculated VAM scores during the 2012-2013 school year, and afterwards discontinued using VAM for evaluative purposes. The qualitative component was a classroom observation in which principals rate teachers using a standards-based rubric. The rubric was known as the COMPASS Teacher Performance Evaluation Rubric, referred to henceforth as the COMPASS rubric. Although the COMPASS system addressed the selection criteria of RTTT, there were scoring irregularities during the implementation of the system and its individual components that warrant further examination.

## **Statement of the Problem**

The practice of classroom observations for teacher performance evaluations has evolved from emphasizing character traits or behavior profiles, to evaluators using standardized rubrics with multiple domains, rating categories, and detailed descriptions of expected performance levels (Danielson & McGreal, 2000; Ellett & Teddlie, 2003; McGreal, 1983). While the field of teacher evaluation was developing before RTTT, policies and practices varied between states (Darling-Hammond et al., 2012). State systems often produced positively skewed distributions of performance ratings (Danielson & McGreal, 2000). In response to these state issues, federal policymakers used the RTTT competition to influence teacher evaluation policies. The selection criteria of RTTT required evaluation instruments to “differentiate effectiveness using multiple rating categories” (USDOE, 2009, p. 9). The RTTT selection criteria also required state policymakers to use the results of teacher evaluations to “develop teachers ... by providing relevant coaching, induction support, and/or professional development” (USDOE, 2009, p. 9).

With the onset of RTTT, the LDOE implemented COMPASS to evaluate teacher performance during the 2012-2013 school year. In August 2013, LDOE staff released the official final statewide scoring distributions from the implementation of the COMPASS rubric. Compared to observations conducted by principals, scores calculated from SLTs or VAM identified a larger percentage of ineffective teachers (LDOE, 2013a). The 2012-2013 VAM distribution categorized 8% of VAM eligible teachers as Ineffective (LDOE, 2013a). In addition, 3% of all teachers were found to be Ineffective based on their SLTs (LDOE, 2013a). These results were in contrast to the 0.35% of teachers statewide that principals rated as Ineffective on the COMPASS rubric (LDOE, 2013a). Thus, there was a positive skew in scores from the

COMPASS rubric that was consistent with previous results of other evaluation instruments (Cantrell & Kane, 2012; Daley & Kim, 2010; LDOE, 2013a).

Within the appendix of their report, LDOE staff disaggregated the data by district and presented it by rating category. Further data, related to the distribution of scores within districts and the number of teachers rated as ineffective, was not reported beyond the first year of COMPASS implementation (K. Nesmith, personal communication, October, 24, 2014). Certain patterns, specific to rural districts, emerged from the district data. For example, rural school districts reported the lowest results in student achievement and also reported low percentages of ineffective teachers (LDOE, 2013a). In 29 of the 68 school districts in Louisiana, principals did not rate any teacher as Ineffective on the COMPASS rubric (LDOE, 2013a). All of these 29 districts were rural school districts, as defined by a distance greater than five miles from the nearest urban center (Provasnik et al., 2007). Furthermore, many of these districts were found in an area known as the Mississippi Delta, along the northeastern border between Louisiana and Mississippi. This pattern suggested that characteristics of rural districts may have influenced principals' decision making during the classroom observation process. The culture and context of rural schools, as potential factors influencing principals during the rating process, are explored further in this study's literature review.

The COMPASS system and its implementation warrant a thorough examination due to the positive skew in the rating distribution of classroom observations as well as the misalignment between the results of teacher evaluations and measures of student learning in rural districts. This study undertook one aspect of such an examination by seeking the views of rural principals regarding the COMPASS rubric and the factors that influenced their rating of teachers.



### **Rationale for the Study**

The persistence of a positive skew in teacher ratings on the COMPASS rubric and the discrepancy between these ratings and measures of student learning in rural school districts indicate the possible existence of factors influencing the evaluation process. As a response by Louisiana policymakers to the selection criteria of RTTT, COMPASS was designed to discern levels of teacher effectiveness. Thus, there is a need for inquiry regarding the implementation of COMPASS and the ratings given to teachers.

### **Conceptual Framework**

Principals were the focal point of this study, as they conducted classroom observations and rated teachers using the COMPASS rubric. Instead of simply holding Louisiana principals responsible for generating the positive skew in the statewide distributions, this study sought to understand the factors that influenced rural principals in their rating decisions by exploring their experiences during the first year of COMPASS implementation, and to provide an understanding of their scoring decisions and the possible influence of external factors. To understand how social and environmental factors may influence principals, this section examines a conceptual framework informed by the rural culture and context.

Policy implementation does not occur in isolation; all organizations have distinct cultures that influence the behavior of their members (Hatch, 2012; Lindahl, 2006). An understanding of organizational culture provides scholars with insight into how organizational leaders, such as principals, behave. The study of organizational culture has moved beyond acknowledging that culture is an inherent aspect of all organizations to exploring both how organization members influence the culture and how the culture influences organization members (Hatch, 2012). Schein (2004) defined culture as:

A pattern of shared basic assumptions that was learned by a group as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems (p. 17).

In his definition, Schein (2004) explained not only what culture is, but also how it is created and perpetuated. Culture shapes accepted patterns of behavior in an organization (Kilmann, 2011; Lindahl, 2006). Older members, rituals, and symbols of the organization socialize new members to the cultural norms, who then begin to favor thinking in terms of collective norms and behaviors (Schein, 2004; Swidler, 1986). Furthermore, it is often difficult for individuals to distinguish their own identities from that of the dominant cultural narrative, as many aspects of culture are hidden from members (Schein, 2004). Culture determines the collective activity of an organization by influencing the behavior of organizational members (Denison, 1990; Hatch, 2012).

Schools, as organizations, have their own cultures which influence how their individual members respond to internal and external situations (Deal & Peterson, 1999). Culture is important to this study because rural schools contain a distinct culture compared to non-rural schools (Capper, 1993; Lee & McIntire, 1999; Theobald & Nachtigal, 1995). The unique culture of rural schools may influence educators to adopt certain behaviors and values (Lee & McIntire, 1999).

Culture is also distinct from the local context. Culture constitutes the shared social aspects of an organization, while context constitutes the environment in which activity and interactions occur (Johnson, Kraft, & Papay, 2012; Schein, 2004). The context of a rural school often contains factors such as geographical isolation and financial hardship (Capper, 1993; Maranto & Shuls, 2012). The physical environment alone does not determine context. For example, policies in an organization add to the context in which individuals operate and make

decisions (Darling-Hammond, Wise, & Pease, 1983). Context, like culture, influences the decision making and behavior of individuals whose behaviors are often in response to factors in their immediate environment (Johnson et al., 2012). Culture and context exist in a reciprocal relationship. Aspects of context may influence the development of culture, and culture may influence the methods by which individuals address their local context.

In this study, I sought information from principals to analyze how rural school culture and the context of the rural environment influenced the rating of teachers on the COMPASS rubric. The use of culture and context as a conceptual framework will be discussed further in Chapter 2.

### **Purpose Statement**

This study examined the experiences of principals in rural schools and the factors that influenced their rating decisions. Contextual factors of interest to this study included, but were not limited to, scarcity of financial resources, inadequate professional development, and a shortage of certified teachers. The culture of rural schools is also of interest for its potential to influence the behavior of principals. Therefore, the purpose of this study was to explore how principals in rural Louisiana school districts perceived the cultural and contextual circumstances that influenced their rating of teachers on the COMPASS rubric.

### **Research Questions**

This study's research questions drew from the experiences of three principals in rural school districts. The testimonies provided by principals allowed the exploration of the scoring distributions created during the implementation of COMPASS. Three questions guided this study.

How do Louisiana principals in rural school districts describe:

1. The influence of the rural culture and context on their scoring decisions?
2. The procedures used by the Louisiana Department of Education to prepare and support evaluators during the implementation of COMPASS?
3. Their reasoning of whether or not they rated teachers as Ineffective?

### **Definition of Terms**

The following terms and their respective definitions are crucial to the understanding, direction, and purpose of this study.

*COMPASS system.* An evaluation system developed by Louisiana policymakers to comply with the selection criteria for RTTT. For teachers, the COMPASS system includes annual evaluations with a final score dependent on the ratings received on the COMPASS rubric and the academic growth of the students they teach (LDOE, 2013a).

*COMPASS Rubric.* A five item instrument with four rating categories which principals used to assess the job performance of teachers in the context of effective classroom instruction (Danielson, 2011a; LDOE, 2013a).

*Principal.* The lead administrator of a school. The principal evaluates teachers using the COMPASS rubric.

*Rural Schools.* A rural school is one that is of a distance greater than five miles from an urban center (Provasnik et al., 2007). While rural areas are defined by their distance from urban centers, urban centers are defined by their populations. The state of Louisiana has seven urban population centers, consisting of 100,000 or more residents: Shreveport, Monroe, Alexandria, Lafayette, Lake Charles, Baton Rouge, and New Orleans (Provasnik et al., 2007).

## **Delimitations and Limitations**

### **Delimitations**

This study was delimited to principals in rural Louisiana public school districts. Selected principals must have evaluated teachers using the COMPASS rubric during the 2012-2013 school year. This study was delimited further to principals within districts wherein students performed below the state average in academic achievement during the 2012-2013 school year, and who did not rate any teachers as Ineffective. The intent of this delimitation was to focus on specific cases where measures of student achievement were not in alignment with COMPASS scores. Excluded from the study were principals of private schools and principals of schools that did not implement the COMPASS rubric.

### **Limitations**

Several factors limited the findings of this study. First, by utilizing principals as participants, this study may have been limited by factors related to their recollections of classroom observations. The recollection and interpretation of experiences by principals during the interview process may have resulted in discrepancies between their experiences and testimony. Second, time was a limiting factor because this study asked principals to recall their experiences from the 2012-2013 school year. If the contextual and cultural factors that influenced their decisions changed during the time interval between their use of the COMPASS rubric and the interview, then their perspectives may have also changed. Third, the study was limited by its reliance on the COMPASS rubric. Fourth, the study was limited by the perspectives of the principals, since they may have had differing approaches to rating teachers, even on the same rubric (Wickert, 1987). Finally, the study was limited by the culture and context that it sought to examine, since the culture and context of the rural school may have

altered how principals responded. For example, the principals may have faced cultural pressure to hide information that did not reflect positively on their rural school or community. The rural context may have also influenced principals to shift their responses to environmental concerns that were a greater priority than accurately rating a teacher. For example, given the high attrition of teachers in rural districts, principals may have inflated performance ratings as a means to retain teachers.

### **Position of the Researcher**

I grew up in the rural community of Patouville, which is in the southern portion of Iberia Parish in south Louisiana. After college graduation, I taught for three years predating the COMPASS evaluation system as a high school science teacher in a Louisiana public school. I am currently the Executive Director of the Associated Professional Educators of Louisiana (A+PEL), a non-union teacher association with chapters throughout the state. Prior to my appointment to this position in January 2014, I was a staff member of A+PEL for three years. As a staff member, I worked on a project in which I presented the COMPASS system to various stakeholders such as teachers and principals. In my current position I am in direct contact with LDOE and BESE staff. I recognized the potential for bias during the research process and addressed this issue through specific methods that are detailed in the methodology chapter.

### **Chapter Summary**

This chapter defined the focus of this study: to examine the experiences of principals and the factors they considered in the ratings of classroom teachers under the COMPASS system. Historical context was given for the emergence of the COMPASS system from the RTTT policy initiative. Although policymakers intended COMPASS to differentiate between levels of teacher job performance, statewide data showed the existence of a positive skew among teacher

performance ratings. This skew was particularly evident in several rural parishes, in which principals did not rate any teachers as Ineffective.

Since principals evaluated teachers using the COMPASS rubric, their perspectives regarding the rating process are essential to understanding the rating distribution. This chapter framed the results generated by principals and the possible connection between those results and the rural culture and context. Since cultural and contextual factors may have influenced principals, this study explored the ratings decisions of principals in rural school districts during the first year of COMPASS implementation.

## CHAPTER 2: LITERATURE REVIEW

### Introduction

This study investigated the influential role of culture and context on principals in their rating decisions on the COMPASS rubric, which the Louisiana Department of Education (LDOE) implemented in the 2012-2013 school year. Louisiana policymakers developed the COMPASS system of teacher evaluation to compete for federal funding under the Race to the Top (RTTT) competition (Hallgren, James-Burdumy, & Perez-Johnson, 2014). As part of the COMPASS system, principals were responsible for evaluating teachers using a standards-based rubric, which policymakers derived from the *Danielson Framework for Effective Teaching*. The results from the first year of COMPASS usage were publically reported by the LDOE at a meeting of the Board of Elementary and Secondary Education (BESE) in the fall of 2013.

State officials reported a positive skew in the data from the first full year of usage of the COMPASS rubric. Based on data compiled from the final reports from districts during the 2012-2013 school year and reported in September 2013, approximately one-third of one percent (0.35%) of teachers received an Ineffective rating, 9% of teachers received an Effective Emerging rating, 63% of teachers received an Effective Proficient rating, and 27% of teachers received Highly Effective ratings on the COMPASS rubric (LDOE, 2013a). The distribution of teachers rated in each rating category on each performance item of the COMPASS rubric is displayed in Table 1 below.



Table 1: 2012-2013 Statewide Percentages of Teachers Scoring in Each Rating Category Across the Items on the COMPASS Teacher Performance Evaluation Rubric

Teachers	Average Score 1.0 – 4.0	Ineffective 1.0 - 1.5	Effective: Emerging 1.5 - 2.5	Effective: Proficient 2.5 - 3.5	Highly Effective 3.5 - 4.0
1c: Setting instructional outcomes	3.3	0.51%	6%	43%	51%
2c: Managing classroom procedures	3.2	0.49%	6%	46%	48%
3b: Using questioning and discussion techniques	3.0	0.78%	10%	59%	31%
3c: Engaging students in learning	3.1	0.55%	7%	56%	37%
3d: Using assessment in instruction	3.1	0.67%	8%	52%	40%

Source: Louisiana Department of Education (2013a)

These results were consistent with the trend reported from preliminary data at the March 2013 BESE meeting, at which the LDOE indicated that the percentage of teachers expected to get an Ineffective rating was less than 1% (LDOE, 2013b).

Under COMPASS policy, principals were responsible for scoring teachers during classroom observations. Ho and Kane (2013) conducted research on the scoring decisions of principals and took note of certain patterns in the rating decisions of principals. First, they observed that principals tend to avoid scoring teachers on the lowest or highest categories on performance rubrics (Ho & Kane, 2013). Second, they recorded principals as giving higher scores to their teachers when compared to external evaluators or even teachers observing their peers (Ho & Kane, 2013), however, additional factors may have influenced principals to maintain a positive skew in the distribution of observational ratings. Furthermore, the skew found in the results from the COMPASS rubric exceeded the descriptions of scoring issues noted by Ho and Kane (2013).

The results of the COMPASS rubric were often in conflict with the reported results for the academic achievement and progress of Louisiana school districts (LDOE, 2013a). In their

first-year COMPASS report, state policymakers juxtaposed the percentage of students passing grade level examinations with the results of teacher evaluations (LDOE, 2013a). For example, the Franklin Parish School District had 57% of students performing at grade level compared to a statewide average of 68%; however, Franklin Parish principals rated 38% of their teachers as Highly Effective (LDOE, 2013a). Furthermore, Franklin Parish principals did not rate any teachers as Ineffective during their classroom observations (LDOE, 2013a). This was but one example of a school district's disparity between the district's academic performance and the rating decisions of principals.

Furthermore, many of the rural districts were at the bottom of state rankings on both student growth and achievement, and the scoring results were positively skewed (LDOE, 2013a). For example, Union Parish student proficiency scores on state standardized exams rank at the 9<sup>th</sup> percentile statewide, yet principals did not rate any teachers as Ineffective, and one out of every five teachers was rated as Highly Effective (LDOE, 2013a). Many of the districts with lower than average student performance but positively skewed teacher evaluation scores were rural; additional issues, such as the culture of the schools, difficulty in retaining teachers, and availability of support for principals, may have influenced the rating decisions of principals. In total, principals in 13 school districts with student proficiency levels below the state average did not rate a single teacher in their district as Ineffective (LDOE, 2013a).

In this literature review, I address four major areas to understand the factors that may have influenced rural Louisiana principals' evaluations of teachers using the COMPASS rubric during the first year of its implementation. These areas are teacher evaluation systems, teacher evaluation in Louisiana, the implementation of COMPASS, and a conceptual framework of culture and context associated with rural schools.

## **Teacher Evaluation Systems**

I first explore the development of teacher evaluation systems in a historical context leading to an analysis of standards-based rubrics. Then, current frameworks and systems for teacher evaluation are examined. I describe nationally recognized systems, current trends in individual states with policies aligned to RTTT, and teacher evaluation policies in other countries.

### **General History of Teacher Evaluations**

Classroom observations by principals historically served as a common means for the evaluation of teacher job performance (Daley & Kim, 2010). Before the 1980s, local administrators developed their own teacher evaluation instruments based on traits instead of teaching practices (McGreal, 1983). For example, principals associated effective teachers with social standing in the community, morality, and personality traits (Ellett & Teddlie, 2003). At the beginning of the 1980s, rubrics became more prominent in the field of teacher evaluation (Ellett & Teddlie, 2003). These early rubrics did not define standards of practice but instead charged principals with subjectively comparing teachers (McGreal, 1983). Critics who described the pre-standards era noted a difficulty in distinguishing between levels of teacher performance or student learning from subjective judgments (Cantrell & Kane, 2012; Daley & Kim, 2010; Shakman et al., 2012). These teacher evaluation practices did not associate evaluation results with systems of support or consequences. For example, evaluation systems in the 1980s did not require conferences between teachers and principals nor the revocation of the certifications of ineffective teachers (McGreal, 1983).

Before standards-based rubrics, policymakers and researchers designed teacher evaluation rubrics with categorical ratings but did not provide descriptions or criteria to assist

evaluators in their rating decisions (Danielson & McGreal, 2000; McGreal, 1983). During both the 1980s and 1990s, teacher evaluation systems consistently categorized the majority of teachers as scoring in the highest performance categories (Cantrell & Kane, 2012; Daley & Kim, 2010; Danielson & McGreal, 2000; Shakman et al., 2012). Danielson and McGreal (2000) described this era of teacher evaluation practices as predominantly:

based on outmoded criteria, observations are conducted on the run by poorly-trained evaluators who are not sure what they should be looking for, and virtually all teachers are rated at the top of whatever scale is used. That is, the way evaluations are conducted in the vast majority of districts, they serve neither of the functions for which they are intended, ensuring quality and promoting professional learning (p. 2)

There were isolated cases of reform in the field of teacher evaluation during the 1980s as individual school systems began to implement accountability policies. Policymakers, such as those in Salt Lake City (Wise, 1984), prescribed remedial job training for those teachers who failed to perform to basic standards of professional practice. In other early initiatives policymakers increased the number of observations conducted and required conferences as a form of coaching and assistance (McGreal, 1983; Weiss & Weiss, 1998; Wise, 1984). However, in the early 1980s, policymakers did not adopt these efforts on a statewide level (Wise, 1984).

During the late 1980s and early 1990s, policymakers began to develop rubrics and accountability policies on a statewide level (Ellett & Teddlie, 2003; Johnson Jr., 1999). The rubrics of the late 1980s were compliance-based and did not distinguish between specific levels of performance (Ellett & Teddlie, 2003; Shinkfield & Stufflebeam, 1995). Evaluators often rated teachers in a dichotomous paradigm of acceptable or unacceptable performance on a limited number of performance items (Shinkfield & Stufflebeam, 1995). Compliance-based rubrics focused on the behavior of teachers and did not hold teachers responsible for the behavior of students (Ellett & Teddlie, 2003) or for measures of student learning (Shinkfield & Stufflebeam,

1995). Practitioners tried to improve rubrics by including standards of teacher practice that researchers positively associated with student learning (Shinkfield & Stufflebeam, 1995). Policymakers established higher standards for instructional practice (Shinkfield & Stufflebeam, 1995). Some districts, such as Rochester, New York, and Cincinnati, Ohio, also began providing incentives, such as merit pay, for high performing teachers (Weiss & Weiss, 1998). Governor Roemer attempted similar efforts in Louisiana in the early 1990s (Johnson Jr., 1999).

By the late 1990s, policymakers aligned teacher evaluation with systems for professional growth, and it became a common policy method to improve classroom practices (Weiss & Weiss, 1998). Policymakers successfully implemented teacher evaluation policies on a statewide level. For example, in 1998, policymakers required all new teachers in California to receive an evaluation of their classroom practices (Pecheone & Chung, 2006). Teacher evaluation policies came under additional scrutiny as policymakers began to hold teachers accountable for student learning.

Research on the impact of effective teaching on student learning brought greater attention to the field of teacher evaluation. After analyzing possible factors affecting student learning, Sanders et al. (1997) determined that the teacher was the primary school-based factor influencing student learning. Researchers and policymakers therefore began to view effective teachers as those educators who were able to produce higher than expected gains in student learning (Goe, Bell, & Little, 2008; Ingvarson & Rowe, 2008). Advocates of teacher evaluation continued to maintain the importance of effective teaching on increasing student learning. Cantell and Kane (2012) noted, “on average, students of teachers with higher teacher effectiveness estimates outperformed students of teachers with lower teacher effectiveness estimates” (p. 8); Cantrell

and Kane (2012) based these estimates of effectiveness on how student test scores improved in previous years.

During the 2000s, policymakers began to use teacher evaluations systems to address issues of job performance among teachers as a means to exert a positive influence on student learning. Because of the precedent of positively skewed results, critics viewed local and state systems as unable to discern between levels of job performance (Daley & Kim, 2010; Danielson & McGreal, 2000). In response to these criticisms, policymakers sought to establish uniform standards of practice for all teacher evaluation systems. RTTT represented a federal intervention to systematically improve teacher evaluation systems and to utilize the evaluation results to address the job performance of teachers (Cantrell & Kane, 2012). The selection criteria for one of RTTT's major reform areas, Great Teachers and Leaders, specifically charged state policymakers to develop evaluation systems that:

inform decisions regarding (a) Developing teachers and principals, including by providing relevant coaching, induction support, and/or professional development; (b) Compensating, promoting, and retaining teachers and principals, including by providing opportunities for highly effective teachers and principals to obtain additional compensation and be given additional responsibilities; (c) Whether to grant tenure and/or full certification to teachers and principals using rigorous standards and streamlined, transparent, and fair procedures; and (d) Removing ineffective tenured and untenured teachers and principals after they had ample opportunities to improve, and ensuring that such decisions were made using rigorous standards and streamlined, transparent, and fair procedures (USDOE, 2009, p. 9).

These selection criteria recommend consequences such as tying licensure and tenure to teacher performance; these consequences were controversial when implemented at the state or local level during the 1980s and 1990s (Johnson Jr., 1999; Popham, 1972). These policies, with their direct consequences on the job security and livelihood of teachers, often faced criticism (Johnson Jr., 1999). This criticism was often because educators perceived that the measures used to establish teacher effectiveness were unreliable or invalid (Johnson Jr., 1999). In response to

both the selection criteria of RTTT and the concerns of educators, state policymakers sought to develop valid and reliable means of rating the job performance of teachers.

**Standards-based rubrics.** Although the RTTT selection criteria did not describe an effective teacher evaluation system, the selection criteria required states to adhere to certain guidelines. The RTTT selection criteria specifically required state policymakers to:

Design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that (a) differentiate effectiveness using multiple rating categories that take into account data on student growth as a significant factor, and (b) are designed and developed with teacher and principal involvement (USDOE, 2009, p. 9).

These requirements represent previously described developments in the field of teacher evaluation. The selection criteria required states to incorporate student learning in their evaluation systems and also required policymakers to differentiate of teacher performance through the use of multiple rating categories. However, many pre-RTTT systems contained multiple categories and these systems still produced results that were positively skewed (Danielson & McGreal, 2000; Shakman et al., 2012). There was no correlation between the number of rating categories and the ability of a system to discern performance levels.

Although RTTT did not require classroom observations as a measure of teacher effectiveness, policymakers in participating states continued observations to measure teacher performance (Shakman et al., 2012). A survey of state practices after the RTTT competition revealed that states were either adopting or modifying established standards-based rubrics from Charlotte Danielson's *Framework for Effective Teaching* or from the National Institute for Effective Teaching (NIET) (Cantrell & Kane, 2012).

Ingvarson and Rowe (2008) described the importance of performance standards for ensuring teacher quality and recommended the use of standards-based rubrics for performance evaluations. Standards-based rubrics, as defined in Chapter 1, provided evaluators with

descriptions of each rating category for every performance item (Milanowski et al., 2004; Odden, 2004). The COMPASS rubric was a standards-based rubric adopted by LDOE policymakers in response to RTTT. This study analyzed aspects of effective teacher evaluation measures to better understand the design of the COMPASS rubric and its alignment with research-based practices.

**Validity and reliability of standards-based rubrics.** Federal policymakers did not include criteria related to validity or reliability of evaluation systems within the RTTT selection criteria. However, researchers described validity and reliability as important aspects of any measure of teacher effectiveness (Cantrell & Kane, 2012). The RTTT selection criteria did require measures of teacher effectiveness to be associated with student learning as a valid and reliable measure would ensure that ratings on performance indicators positively correlated with gains in student learning. Gallagher (2004) tested the findings of subject-specific standards-based rubric and the resulting teacher evaluation scores as valid predictors of student learning and found strong and positive correlations between teacher evaluation and student achievement. Gallagher's (2004) research, however, did not align with the findings of other research in the field of teacher evaluation. In a survey of research literature on standards-based rubrics the highest correlation coefficient was .50 (Milanowski et al., 2004). These coefficients represented a low to moderate correlation between scores generated from standards-based rubrics and student learning (Milanowski et al., 2004). Kimball et al. (2004) also found weak correlations between teacher evaluation scores on a shortened version of the Danielson rubric containing seven components and student achievement on the Terra Nova exam series in the third, fourth, and fifth grades. However, Kimball et al. (2004) could not establish statistical significance for their findings in the third grade.



Kimball and Milanowski (2009) noted that reliability between evaluators was a serious issue in the field; they examined the variances in scoring between evaluators and their results were inconclusive. Cantrell and Kane (2012) also tested reliability between evaluators and compared the rating decisions of principals to external evaluators. They noted the following when they compared principal rating decisions to those of others:

Although administrators gave higher scores to their own teachers, their rankings of their own teachers were similar to those produced by peer observers and administrators from other schools. This implies that administrators are seeing the same things in the videos that others do, and they are not being swayed by personal biases (p. 18-19).

This was a key finding, because it revealed that a standards-based rubric, with descriptors to guide evaluators, contains a level of reliability among raters. This inter-rater reliability even extended to evaluators who were external to the school site.

Researchers supported by the Bill and Melinda Gates Foundation sought to reduce the variation between the scoring decisions of different evaluators. Cantrell and Kane (2012) identified specific practices that increased the reliability of classroom observations. In studying the use of a standards-based rubric which consisted of ten performance items from the Danielson rubric, they found that using the numerical average of multiple observations scored by multiple evaluators increased the reliability to the “extent to which the variation in results reflects consistent aspects of a teacher’s practice, as opposed to other factors such as differing observer judgments” (p. 18). Cantrell and Kane (2012) found that the reliability of rating decisions increased with each additional incident of observation. Through the COMPASS system, Louisiana policymakers required principals to conduct at least two observations.

Previous teacher evaluation systems and their associated classroom observation rubrics were unable to distinguish between levels of job performance, correlate their results to student learning, or address criticisms regarding their validity and reliability. However, after the RTTT

competition, policymakers adopted standards-based rubrics to measure teacher performance, which provided a description for each performance item and each rating category. These descriptions provided guidance for the scoring decisions of evaluators. Researchers have positively correlated the results of standards-based rubrics to gains in student achievement, but these correlations were mostly limited to statistically weak correlations. Similar research has not been conducted on the COMPASS rubric, however. While this section explored components and practices of standards-based rubrics, the next section examines specific rubrics and evaluation systems.

### **Current Teacher Evaluation**

In this section, I examine several teacher evaluation systems and their associated instruments. All of the selected systems were recognized within the research field or widely utilized by evaluators. I explore teacher the evaluation systems within three areas: independent systems prior to RTTT, the systems of individual states after RTTT, and selected examples from other nations.

**National Board for Professional Teaching Standards.** The National Board for Professional Teaching Standards (NBPTS) evaluated teachers who were pursuing National Board Certification (NBC). Smith, Gordon, Colby, and Wang (2005) recognized efforts by the NBPTS, which began in 1987, as one of the earliest national attempts to identify effective teachers who positively influence student learning and achievement. The NBPTS used an evaluative process consisting of a standards-based system with multiple measures of teacher effectiveness (Cavalluzzo, 2004). Cavalluzzo (2004) described the evaluation process:

Applicants for NBC must prepare a portfolio with three classroom-based entries and a combined Document Accomplishment Entry that describes work with families and caregivers, as well as participation in the professional community. In addition, each

applicant must complete six 30-minute exercises at the National Board's Assessment Center (p. 6).

Therefore, under the NBC process, evaluators evaluated teachers on both a classroom observation and a series of assessments. A committee of assessors trained by the NBPTS evaluate videos of classroom teaching as part of the certification process (Cavalluzzo, 2004). The assessments covered both content and relevant pedagogical knowledge (Berry, 2002). Smith et al. (2005) explained the importance of examining the content knowledge of teachers, because "content knowledge and instruction were inextricably related to students' needs and students' learning" (p. 9). Throughout the entire process, the NBPTS evaluators assessed teachers on five principles. These principles were:

Teachers are committed to students and their learning. Teachers know the subjects they teach and how to teach those subjects to students. Teachers are responsible for managing and monitoring student learning. Teachers think systematically about their practice and learn from experience. Teachers are members of learning communities. (Darling-Hammond, 1999, p. 6).

Researchers have highlighted the NBPTS for its focus on both student learning and teacher knowledge (Darling-Hammond, 1999, 2006; Smith et al., 2005). Furthermore, Cavalluzzo (2004) investigated links between NBC teachers and increases in student learning and concluded that the NBC was an indicator of teacher quality. Cavalluzzo (2004) found that:

In particular, NBC proved to be both an effective signal of teacher quality and a valid discrimination of teacher quality among applicants. Indeed, seven of nine indicators of teacher quality that were included in the analyses resulted in appropriately signed and statistically significant evidence of their influence on student outcomes. Among those indicators, having an in-subject-area teacher, NBC, and regular state certification had the largest effect sizes (p. 34).

Smith et al. (2005) also conducted research to examine the link between NBC and student writing samples. From an examination of student writing samples, Smith et al. (2005) found that students with NBC teachers had a higher quality of writing than students who had teachers

without NBC. Similarly, Cavalluzzo (2004) concluded that NBC was an indicator of teacher effectiveness. The NBPTS represented an effective system of teacher evaluation as those teachers who were rated highly enough to earn NBC were positively associated with gains in student learning.

**Interstate New Teacher Assessment and Support Consortium (InTASC).** As an early effort to improve teacher quality, the NBC process influenced practitioners in the development and selection of professional standards in other systems, such as the Interstate New Teacher Assessment and Support Consortium (INTASC) (Cavalluzzo, 2004; Darling-Hammond, 1999; Gordon, 2002). Whereas the NBC process was an advanced, voluntary certification for experienced teachers, the INTASC standards were managed by the Council of Chief State School Officers (CCSSO) and associated with pre-service teachers and their initial certification (Darling-Hammond, 2006). Annetta and Dotger (2006) described the INTASC as “a collaboration of state and national educational organizations dedicated to reform the preparation, licensing, and professional development of teachers” (p. 42). Cavalluzzo (2004) viewed the INTASC standards as a continuation of the work of the NBPTS.

The latest version of the INTASC standards listed ten standards in each of the four areas listed below:

1. The Learner and Learning: Learner Development, Learning Differences, Learning Environments
2. Content: Content Knowledge, Application of Content
3. Instructional Practice: Assessment, Planning for Instruction, Instructional Strategies
4. Professional Responsibility: Professional Learning and Ethical Practice, Leadership and Collaboration (CCSSO, 2011).

Darling-Hammond (1999) highlighted the performance-based nature of the standards: “they describe what teachers should know, be like, and be able to do, rather than listing courses that teachers should take in order to be awarded a license” (p. 15). The INTASC standards were

designed not as an evaluation system with consequences tied to performance, but as a means to develop teachers (Annetta & Dotger, 2006). The standards provide teachers with opportunities to reflect and seek support (Annetta & Dotger, 2006). Although the INTASC standards would not meet the selection criteria of RTTT, researchers recognized these standards as establishing foundational practices for teacher evaluation and for influencing the preparation of teachers (Annetta & Dotger, 2006; Darling-Hammond, 1999; Ladson-Billings & Darling-Hammond, 2000).

**Danielson's Framework for Effective Teaching.** In 1996, Danielson published the *Framework for Effective Teaching*. This work later developed into an instrument of classroom observation known as the Danielson rubric. Danielson (1996) designed this framework to detail the professional knowledge and skills expected of teachers. The framework contains four domains and 22 components. Domains were the overall area into which components were categorized. Each domain contains a number of components that described the actions, skills, and knowledge expected of teachers (Danielson, 2007). In designing the initial framework, Danielson (2007) aligned the domains and components with the INTASC standards. The framework's domains and components have remained consistent since Danielson developed it. (Danielson, 1996, 2007, 2011b).

The domains and components of the Danielson framework are as follows:

1. Planning and Preparation: Demonstrating Knowledge of Content, Demonstrating Knowledge of Students, Setting Instructional Outcomes, Demonstrating Knowledge of Resources, Designing Coherent Instruction, Designing Student Assessments
2. The Classroom Environment: Creating an Environment of Respect and Rapport, Establishing a Culture for Learning, Managing Classroom Procedures, Managing Student Behavior, Organizing Physical Space
3. Instruction: Communicating with Students, Using Questioning and Discussion Techniques, Engaging Students in Learning, Using Assessment in Instruction, Demonstrating Flexibility and Responsiveness

4. Professional Responsibilities: Reflecting on Teaching, Maintaining Accurate Records, Communicating with Families, Participating in a Professional Community, Growing and Developing Professionally, Showing Professionalism (Danielson, 2007, pp. 3-4).

Initial versions of the framework (Danielson, 1996, 2007) described components through elements, which served as evidence of a teacher addressing a component. For example, under the domain of instruction the component of setting instructional outcomes was associated with the elements of “value, sequence, alignment, clarity, balance, and suitability for diverse learners” (Danielson, 2007, p. 3).

Danielson (1996, 2007) initially designed the framework as a professional development tool to address the needs of teachers through a system of classroom observation, feedback, and support. As previously noted, after RTTT, policymakers in several states used all or portions of the Danielson framework as a rubric for classroom observation. In response to its inclusion in teacher evaluation systems, Danielson and her fellow researchers released a new guide for the framework. This publication modified the framework to further its use as a rubric for teacher evaluation (Danielson, 2011b). This update included the development of the component elements into detailed descriptions of performance (Danielson, 2011b). Danielson (2007) tied these descriptions to specific rating categories for each component: unsatisfactory, basic, proficient, and distinguished.

The Danielson framework has seen widespread use across the nation, including use as a rubric for teacher evaluation, both before and after RTTT (Clayton, 2013; Heneman III, Milanowski, Kimball, & Odden, 2006; Kellor, 2005). Danielson (2007) described the framework as “grounded in a body of research that seeks to identify principles of effective practice and classroom organization” (p. 21). Researchers have supported the Danielson rubric as a valid and reliable measure of teacher effectiveness (Kane & Staiger, 2012; Sartain, Stoelinga, & Brown,

2011). Policymakers in many states and districts used modified versions of the Danielson rubric in their teacher evaluation systems, often reducing the number of performance items (Ho & Kane, 2013; Kimball, White, Milanowski, & Borman, 2004). For example, Louisiana policymakers developed the COMPASS rubric from five performance items found in the Danielson rubric. At the time of this study, neither policymakers nor researchers have compared the validity and reliability of COMPASS to the original Danielson rubric.

**Teacher Advancement Program.** In 1999, the National Institute for Excellence in Teaching (NIET), with financial support from the Milken foundation, created the Teacher Advancement Program (TAP) (Glazerman, McKie, & Carey, 2009; Podgursky & Springer, 2007; Toch, 2008). TAP was a program for school reform (Schacter & Thum, 2005) that focused on evaluating, supporting, and developing teachers (Podgursky & Springer, 2007). Unlike the Danielson framework, TAP contains a comprehensive system of rewards and consequences associated with performance ratings. Through TAP, policymakers provided teachers with leadership opportunities, professional development, and increased compensation (Glazerman et al., 2009). Since TAP's development, school leaders have adopted the system in hundreds of US schools (Podgursky & Springer, 2007).

TAP's evaluation system relied on a rubric known as the Skills, Knowledge, and Responsibilities (SKR) rubric (Whiteman, Shi, & Plucker, 2011). Researchers developed SKR as a standards-based rubric with five rating categories (Whiteman et al., 2011). As Toch (2008) explained:

TAP's modified version of Danielson's teaching standards had three main categories: designing and planning instruction, the learning environment and instruction, and 19 subgroups that target areas such as the frequency and quality of classroom questions and whether teachers are teaching students such higher level thinking skills as drawing conclusions (p. 2).

The performance ratings derived from the TAP rubric were tied to the other components of the program (Podgursky & Springer, 2007). For example, school administrators rewarded teachers who received high performance ratings with additional compensation and opportunities (Glazerman et al., 2009). These additional opportunities may include leadership activities such as coaching, planning lessons, and mentoring of other teachers (Schacter & Thum, 2005). Glazerman et al. (2009) provided further descriptions of the incentives available to teachers: “under the TAP model, teachers can earn extra pay and responsibilities through promotion to Mentor or Master Teacher and can earn annual performance bonuses based on a combination of their value added to student achievement and observed performance in the classroom” (p. 1). For school leaders within a TAP school, teacher evaluation serves as the basis for decisions regarding TAP’s career ladder and compensation system.

Because of the importance of evaluation in the TAP system, NIET researchers seek to continually improve the system’s evaluation practices (Jerald & Van Hook, 2011). Currently, evaluators in TAP schools use “multiple trained observers, including principals and Master and Mentor teachers” (Barnett, Wills, & Kirby, 2014, p. 7). Evaluators in a TAP school received training from NIET on conducting observations, rating teachers, and providing feedback in post-observation conferences (Jerald & Van Hook, 2011). Jerald and Van Hook (2011) examined the results of TAP evaluations for their validity and reliability and found that TAP evaluations “can indeed produce valid, non-inflated ratings, even when such individuals work in the same schools as their evaluators” (p. 4).

Early research (Schacter & Thum, 2005) on TAP schools found greater growth in student scores compared to similar schools that do not follow the TAP system. Further research (Solmon, White, Cohen, & Woo, 2007) encompassing 610 teachers in TAP schools found increased



student growth in comparison to teachers in a control sample consisting of non-TAP schools. Beginning in 2007, leaders in the Chicago Public School (CPS) system implemented TAP (Glazerman et al., 2009). Research (Glazerman et al., 2009) compared TAP schools in Chicago to a control group of schools with similar student populations. The comprehensive CPS study (Glazerman et al., 2009) included data from teacher interviews, student assessments, and school performance in its analysis. The results of the first year of implementation of TAP in Chicago were mixed. Teacher retention increased and teachers reported increased levels of support, but student scores and teacher satisfaction were not significantly different in TAP schools compared to the control group (Glazerman et al., 2009). Furthermore, regarding the culture of the schools and the attitudes of the teachers, Glazerman et al. (2009) established “no statistically significant differences between TAP and non-TAP schools” (p. 33).

Out of approximately 1,400 public schools in Louisiana, 66 schools participated in TAP from 2010 to 2013 (Barnett et al., 2014). When compared to similar schools that did not participating in TAP, these 66 TAP schools demonstrated “significantly increased achievement growth” (Barnett et al., 2014, p. 13). As a comprehensive system, TAP met many of the selection criteria of RTTT. For example, TAP uses multiple rating categories and ties consequences to performance ratings. However, TAP evaluators relied on the SKR rubric as the primary measure of teacher evaluation, and the research was inconclusive regarding how performance ratings correlated with increases in student learning.

**Teacher Evaluation in the United States.** RTTT served as an impetus for states to reform their teacher evaluation policies (Shakman et al., 2012). Before the federal intervention of RTTT, state policy on teacher evaluation varied. For example, only 21 states required teachers to be evaluated annually and only 20 states had teacher evaluation rubrics based on performance

standards (Hazi & Rucinski, 2009). The federal intervention in state education policy by means of RTTT created major changes in state policy across the nation.

An examination of teacher evaluation policy after RTTT identified several trends. First, policymakers made efforts to incorporate measures of student learning into the teacher evaluation process, however, only Colorado, Georgia, New York, and Massachusetts policymakers established processes to research the validity and reliability of their measures in predicting gains in student achievement (Herlihy et al., 2013). A recent survey found that the implementation of RTTT policy varied based on the selection criterion (Hallgren et al., 2014). For example, 30 states used multiple measures of teacher evaluation, but policymakers in only 11 states used the results of their teacher evaluation systems to award teachers through further compensation. While some state policymakers did not implement systems that completely adhered to the selection criteria of RTTT, major shifts in teacher evaluation policy occurred.

Shakman et al. (2012) surveyed state policy changes in response to RTTT and identified four state systems that met the selection criteria and were implemented by the 2010-2011 school year: Delaware, Georgia, North Carolina, and Tennessee. The participating states shared many common features in their teacher evaluation systems including multiple measures of teacher effectiveness, multiple rating categories, and annual evaluations (Shakman et al., 2012). Shakman et al. (2012) found variations in the performance standards of each state's teacher evaluation rubric and the number of rating categories.

Delaware. The federal government awarded funding to the states of Delaware and Tennessee in the first round of the RTTT competition (Hamilton, 2010). Delaware policymakers implemented a statewide teacher evaluation system in 2008 and required annual evaluation of new teachers (Shakman et al., 2012). The Delaware system contains two measures: a reflective

survey taken by the teacher and an evaluation by the principal scored from classroom observations and conferences on student performance (Shakman et al., 2012). Staff from the Delaware Department of Education trained principals on a rubric consisting of five components: planning and preparation, classroom environment, instruction, professional responsibilities, and student improvement (Shakman et al., 2012). Unlike standards-based rubrics, the Delaware rubric only provided indicators of performance and evaluators rate teachers as satisfactory or unsatisfactory (Shakman et al., 2012).

Georgia. Teacher evaluation in Georgia involved a sequence of measures. First, teachers completed a reflection, set goals for student learning, and created a professional development plan (Shakman et al., 2012). Then, principals conducted classroom observations and collected evidence of student performance (Shakman et al., 2012), which may include student work samples. Five components guided this process: curriculum and planning, standards-based instruction, assessment of student learning, professionalism, and student achievement (Shakman et al., 2012). Evaluators utilized a standards-based rubric that provided descriptions for four performance ratings.

North Carolina. Principals in North Carolina evaluate teachers on classroom observations, although teachers must submit professional development plans and document reflection on their practice (Shakman et al., 2012). The frequency of observations and document submission was determined by tenure, with non-tenured teachers undergoing the evaluation process annually (Shakman et al., 2012). Evaluations were in accordance with five standards adopted by North Carolina during the development of its evaluation system: teachers demonstrate leadership, teachers establish a respectful environment for a diverse population of students, teachers know the content they teach, teachers facilitate learning for their students, and

teachers reflect on their practice (Shakman et al., 2012). The North Carolina rubric divided performance into five categories, and evaluators were provided with general performance indicators instead of descriptions specific to each rating category (Shakman et al., 2012).

Tennessee. As one of two first round winners in the RTTT competition, Tennessee received 500 million dollars in federal funding (Hamilton, 2010). The Tennessee system was unique in that it began in 1997 and was modified in response to both NCLB and RTTT (Shakman et al., 2012). Evaluators rated teachers on 14 performance indicators and classified teacher performance as one of four rating categories (Shakman et al., 2012). The Tennessee system was also unique in that performance ratings were tied to experience, with experienced teachers requiring higher ratings in order to receive a satisfactory evaluation (Shakman et al., 2012). In response to RTTT, the Tennessee system now requires annual evaluations for all teachers regardless of experience or tenure (Shakman et al., 2012).

Comparison. Each of these four states shares common practices and policies. For example, every state uses classroom observations by principals as a measure of teacher performance. State policy does vary, however, on the frequency of teacher evaluation. Only Tennessee policymakers mandate that all teachers, regardless of experience, must be evaluated each year. State policymakers also favor the use of reflective surveys and professional development plans to involve teachers in the evaluation process and require evaluators to obtain artifacts of student learning. These measures differ from those of the COMPASS system, which does not include reflective surveys or professional development plans as measures of job performance.

**International Trends in Teacher Evaluation.** Examining the teacher evaluation policies of countries with high levels of student achievement can provide a better understanding of the

impacts of these policies. Established international assessments include the Trends in International Math and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), and the Programme for International Student Assessment (PISA). Both TIMSS and PIRLS were developed and managed by the International Association for the Evaluation of International Achievement (IEA). The IEA consists of participating research centers, such as the Lynch School of Education at Boston College, and approximately 50 countries participate in TIMSS and PIRLS (Mullis, Martin, Foy, & Arora, 2012; Mullis, Martin, Gonzalez, & Chrostowski, 2004). PISA was managed by the Organisation for Economic Co-operation and Development (OECD) and supported by 70 countries as a means of internationally comparing student achievement (Schleicher, 2009). Due to the higher number of countries participating in PISA than TIMSS or PIRLS, this study focused on the results of PISA. PISA was conducted every three years and focused on student achievement in certain subjects (Ho, 2003). The first three PISA studies focused individually on literacy, mathematics, and science (Ho, 2003; Simola, 2005). However, the 2009 and 2012 versions examined achievement in all three subjects (Stacey, 2012). In Asia, students in countries such as the Republic of Korea, Japan, and the area of Hong Kong consistently outperformed US students in all subject areas (Fleischman et al., 2010). From Europe, students in countries such as Finland, Switzerland, Germany, and Denmark outranked US students (Fleischman et al., 2010). Researchers have called attention to the nations of Finland and Korea, where students have consistently held high PISA scores in all three content areas (Fleischman et al., 2010; Kupiainen, Hautamäki, & Karjalainen, 2009; Simola, 2005).

Finland. Simola (2005) recognized Finland for both its high levels of academic achievement and equity in achievement between different socio-economic classes. These

educational successes came from a long-term effort to improve education by the Finnish government. In the 1970s, Finnish policymakers passed the first set of reforms aimed at improving student performance (Kupiainen et al., 2009). These initial reforms centralized academic standards through the adoption of a national curriculum (Kupiainen et al., 2009). Later, during the 1980s, control was split between the National Board of Education, which developed core curriculum and learning objectives and the local cities, which were allowed to edit and develop their local curriculum (Kupiainen et al., 2009).

Finnish policymakers placed responsibility for evaluations on local administrators (Kupiainen et al., 2009). Finnish policymakers may not have felt the same political pressure to create and enforce accountability measures as their American counterparts. The general public expressed in opinion polls that Finnish teachers have high levels of social status and are trusted professionals (Simola, 2005). Furthermore, Finnish college students expressed favorable views for profession and enter into teaching at a high rate, similar to high-status career fields such as law and medicine (Simola, 2005).

South Korea. Students in the Republic of Korea, also known as South Korea, have consistently demonstrated high levels of achievement on the PISA (Fleischman et al., 2010), especially in mathematics (Kang & Hong, 2008). Kang and Hong (2008) compared teacher workforce policies in both South Korea and the US and noted two significant differences. First, pre-service teachers in South Korea faced a highly selective process to become teachers, with college entrance, licensure, and job placement exams (Kang & Hong, 2008), whereas, Angrist and Guryan (2004) noted high passage rates on license exams, even among US Education majors with below average college GPAs. Second, teachers in South Korea were socially revered and highly paid, resulting in low rates of teacher attrition (Kang & Hong, 2008). In the US, Darling-

Hammond and Berry (2006) called attention to high rates of teacher attrition which create additional costs for systems that have to continually recruit and train new teachers.

South Korean policymakers recently sought to create a national teacher evaluation system. Before 2010, South Korean principals were solely responsible for observing and evaluating teachers (Seo, 2012). Policymakers discussed altering the existing teacher evaluation policy in the early 2000s (Xue, 2006), but faced criticism from South Korean educators (Seo, 2012). Then, in 2010, policymakers created a national teacher evaluation system that transitioned the responsibility of classroom observations from principals to “multiple evaluations by multiple evaluators” (Seo, 2012, p. 75); known as the Evaluation of Teacher Professional Development (ETPD), the system required the inclusion of student and parent surveys in teacher performance ratings. Throughout its development and implementation, teachers criticized the ETPD, with a majority of teachers describing the system as not contributing to their professional growth (Seo, 2012).

General International Trends. Two general trends emerged from a survey of the international literature on teacher evaluation. First, policymakers in countries with high levels of student achievement have not implemented national policies governing teacher evaluation. For example, policymakers in Japan, Germany, and Finland (Kupiainen et al., 2009) do not rely on teacher evaluations to improve teacher performance and student learning. Second, in nations that established national teacher evaluation policies, policymakers have only recently begun these efforts. For example, policymakers in South Korea (Seo, 2012), China (Liu & Teddlie, 2003, 2005) and the United Kingdom (Reynolds, Muijs, & Treharne, 2003) enacted teacher evaluation policies only within the last decade.

In the literature on nations with high levels of student achievement and educational equity, researchers did not attribute academic achievement to teacher evaluation policies (Seo, 2012; Simola, 2005). Researchers instead described successful nations as having a high-quality, professional, and selective workforce of teachers (Kang & Hong, 2008; Seo, 2012; Simola, 2005). The status of teachers and the strong educational workforce in high-performing countries may provide an explanation for the differences in policies among countries. Some countries addressed teacher quality and development at the pre-service level, through selectivity in the licensure of those attempting to enter the profession. In countries that were unable to attract a high-quality workforce of educators, policymakers may have enacted evaluation policies to address issues of workforce quality in a job-embedded manner.

**Summary of International Trends.** An examination of teacher evaluation policies on an international scale provided an alternative discourse to the theory that teacher evaluation policies were necessary to improve teacher performance. Policymakers in high-performing countries do not use evaluation policies to address the performance of teachers. Through selective admissions, licensure, or culture, policymakers address teacher quality prior to job placement. Contrary to the current dominant policy narrative in the US and the implications within the selection criteria of RTTT, teacher evaluation policy may not be a significant contributor to student achievement.

### **Teacher Evaluation in Louisiana**

In this section, I detail the progression of teacher evaluation in Louisiana. I describe two systems of teacher evaluation used by Louisiana educators prior to COMPASS before examining the early development of COMPASS. During the discussion on the development of COMPASS, I provide two perspectives on its development: COMPASS as a standards-based rubric and COMPASS as a modification of the Danielson Framework.



## **Louisiana Teacher Evaluation Systems Prior to COMPASS**

Prior to RTTT, LDOE policymakers implemented two statewide teacher evaluation systems. First, Louisiana policymakers implemented LATEP with its STAR instrument for classroom observations before replacing it with LCET. LCET relied on a rubric and removed the controversial aspects associated with LATEP. The examination of these previous systems provided a policy background for the design and implementation of the COMPASS system. The structure, scoring, and training provided by policymakers when STAR and LCET were implemented may have influenced the evaluators in the COMPASS system, as many of these principals evaluated teachers under the previous systems. Furthermore, the prior controversies surrounding STAR and LCET may have profoundly influenced the degree to which Louisiana educators accepted the COMPASS system.

**STAR.** The STAR instrument was developed by Chad Ellett of Louisiana State University (LSU) as the primary evaluation instrument in the LATAP system (Teddle, Ellett, & Naik, 1990). Developers (Teddle et al., 1990) designed STAR with four performance dimensions and 22 performance areas described as teaching and learning components:

1. Preparation, Planning, and Evaluation: Goals and Objectives, Teaching Methods and Learning Tasks, Allocated Time and Content Coverage, Aids and Materials, Home Learning, Formal Assessment and Evaluation
2. Classroom and Behavior Management: Time, Classroom Routines, Student Engagement, Managing Task-Related Behavior, Monitoring and Maintaining Student Behavior
3. Learning Environment: Psychosocial Learning Environment, Physical Learning Environment
4. Enhancement of Learning: Lesson and Activities Initiation, Teaching Methods and Learning Tasks, Aids and Materials, Content Accuracy and Emphasis, Thinking Skills, Clarification, Monitoring Learning Tasks and Informal Assessment, Feedback, Oral and Written Communication

A team of researchers from the LDOE and LSU tested each of these performance dimensions and their respective performance indicators for their reliability under multiple

evaluators (Teddle et al., 1990). While the reliability of each individual indicator may vary slightly depending on the evaluator, overall, the research team from LSU (Teddle et al., 1990) found STAR to be a valid and reliable measure of teacher effectiveness. Critics of the system decried the pilot tests as limited and secretive (Baldwin, 1995); however, the STAR instrument was the only instrument in Louisiana to be tested for reliability, since neither LCET nor the COMPASS rubric underwent a validation process.

Unlike more recent rubrics such as COMPASS or the Danielson rubric, STAR did not contain multiple rating categories or descriptors of effective performance; instead, principals had to score teacher performance indicators as either Acceptable or Unacceptable when using STAR (Teddle et al., 1990). After the initial rating by performance item, principals then generated a final score by the number of Acceptable ratings.

Although STAR underwent pilot testing to ensure its reliability and validity, it still faced opposition. Some teachers noted that the rubric itself was generic and evaluators were not trained to apply STAR to certain scenarios, such as the evaluation of a foreign language teacher (Baldwin, 1995). Teachers also demanded inclusion in the development process of a teacher evaluation system (Baldwin, 1995). These controversies soon led to educators engaging in political activism on a statewide level as teachers protested at the state capital and raised legal challenges to the consequences prescribed for poor performance (Baldwin, 1995). Because of the controversy surrounding the LATEP system as a whole and STAR as an evaluation instrument, BESE invalidated the results of the first year of implementation, and the system was placed on hold until a new system could be developed (Johnson Jr., 1999). Although this controversy occurred during the early 1990s, the controversy reveals a history of resistance to teacher evaluation in Louisiana. These points of contention, such as the fairness of the evaluation, the

secrecy of the pilot testing, and the lack of involvement by educators, were important identifiers for potential controversy over current teacher evaluation systems. Furthermore, while resistance to STAR was pronounced and visible, resistance to policy initiatives can take other forms, such as sabotage. If the culture of a district or school contains an institutional memory of the controversy over the scoring results of STAR, that memory may exert an unseen influence on the scoring in the Compass system.

**LCET.** The LDOE created LCET in response to the controversy over LATEP and STAR (Johnson Jr., 1999). Many of the aspects of LCET can be seen as addressing the prior controversies. Unlike LATEP, LCET was only an instrument of teacher evaluation, not a system of incentives and consequences for teacher performance (Baldwin, 1995). This change removed the controversial consequences tied to performance on STAR, and presented LCET as a non-threatening tool for evaluating performance. LCET was also developed by a public committee of educators from all levels of the educational system and presented to BESE (Baldwin, 1995). This process was in direct response to the controversy over the secrecy and lack of educator involvement in the development of STAR. However, researchers did not conduct research on the reliability of LCET between evaluators or the validity of the performance items in identifying effective teaching during its development. LCET contained less performance items for scoring than STAR.

From its development and implementation, LCET saw continual use in Louisiana until the implementation of COMPASS. Current principals may have used LCET prior to their use of the COMPASS rubric. The contents of LCET merit an examination, since the prior use of LCET may influence principals in their decisions on the COMPASS rubric. Furthermore, principals

may have culturally adopted the scoring expectations associated with LCET, creating a hidden influence upon current COMPASS evaluators.

As an instrument to assess teacher performance, LCET contained 19 performance items categorized in seven components under the domains of management and instruction (LDOE, n.d.). The structure of LCET was as follows:

1. Management
  - a. The teacher maintains an environment conducive to learning
    - i. Organizes available space, materials, and/or equipment to facilitate learning
    - ii. Promotes a positive learning climate
  - b. The teacher maximizes the amount of time available for instruction
    - i. Manages routines and transitions in a timely manner
    - ii. Manages and/or adjusts time allotted for planned activities
  - c. The teacher manages learner behavior to provide productive learning opportunities
    - i. Establishes expectations for learner behavior
    - ii. Uses monitoring techniques to facilitate learning
2. Instruction
  - a. The teacher delivers instruction effectively
    - i. Uses techniques which develop lesson objectives
    - ii. Sequences lesson to promote learning
    - iii. Uses available teaching materials and aids to achieve lesson objectives
    - iv. Adjusts lesson when appropriate
  - b. The teacher presents appropriate content
    - i. Presents content at a developmentally appropriate level
    - ii. Presents accurate subject matter
    - iii. Relates relevant examples, unexpected situations, or current events to the content
  - c. The teacher provides opportunities for student involvement in the learning process
    - i. Accommodates individual differences
    - ii. Demonstrates ability to communicate effectively with students
    - iii. Stimulates and encourages higher order thinking at the appropriate developmental levels
    - iv. Encourages student participation
  - d. Student Assessment
    - i. Monitors ongoing performance of students
    - ii. Provides timely feedback to students regarding their progress (LDOE, n.d.).

Principals rated each of these performance items as Unsatisfactory, Needs Improvement, Area of Strength, or Demonstrates Excellence, with each scoring category converted to a numerical rating between 1 and 4. Although LCET was not a standards-based rubric that described each rating category for each performance item, the supporting documentation for LCET provided some scoring guidance (LDOE, n.d.). While not providing support for rating decisions at the level of detail of a standards-based rubric, the LDOE provided support and training to the evaluators using LCET. However, the LDOE may have not addressed the adjustment from LCET to COMPASS in their training and support for implementing COMPASS. The LDOE based the scoring procedures and expectations of LCET on the frequency of observed actions, while the LDOE based the descriptors of rating categories associated with COMPASS on defined actions and behaviors for each performance item. The transition to COMPASS from the scoring paradigm of LCET may have influenced the rating decisions of principals who are now using the COMPASS rubric, especially if the differences between the rubrics were not addressed through training.

### **Development of the COMPASS Teacher Evaluation System**

Principals continued to use LCET until policymakers implemented the COMPASS system. Since LCET may be viewed as a response to STAR's controversy among educators, COMPASS may be viewed as a response to LCET's deficiencies in meeting the selection criteria of RTTT. RTTT required a more comprehensive and systematic approach to teacher evaluation, while LCET was only a rubric and not a statewide teacher evaluation system (USDOE, 2009). For example, LCET did not incorporate student learning into the evaluation of teachers, did not reward effective teachers, or address the job status of ineffective teachers (USDOE, 2009). However, even if LATEP and STAR had remained in place, new policies would have been

required to meet the selection criteria of RTTT. While LATEP did contain consequences and rewards for teachers at various levels of performance, LATEP did not include student learning as a measure of teacher performance and STAR did not contain multiple rating categories for each performance item (USDOE, 2009). Despite over twenty years of efforts in teacher evaluation, changes to teacher evaluation were necessary for the state of Louisiana to be eligible for RTTT funding.

RTTT, as previously noted, was a powerful impetus for state policymakers to develop new policies, especially in the area of teacher evaluation. State governments competed for funding under the RTTT grant competition, which totaled to 4.35 billion dollars (USDOE, 2009). State policymakers implemented changes in education law and policy in order to adhere to the RTTT selection criteria. The United States Department of Education (USDOE) evaluated the applications using a rubric with two to five selection criteria per area of reform. The USDOE assigned a maximum point value to each criterion that ranged from a low of 5 points to a high of 65 points. Variations in the number of criteria per area of reform and the allotment of points per criterion resulted in differences in the maximum number of points that could be awarded for each area of reform. This number ranged from a low of 47 points to a high of 138 points.

The RTTT area of Great Teachers and Leaders contained five selection criteria for a total of 138 points (USDOE, 2009). The selection criteria for this area were “increasing the pathways for certification of new teachers”, “improving teacher effectiveness based on performance”, “ensuring equitable distribution of effective teachers across schools regardless of poverty or racial composition”, and “improving preparation programs, and providing effective support to teachers and principals” (USDOE, 2009, pp. 9-10). At 58 points, the selection criteria of

improving teacher effectiveness based on performance was worth more points than any other selection criteria (USDOE, 2009).

Louisiana policymakers began to develop the COMPASS system in 2010 after the initial application period for RTTT funding. LDOE policymakers and staff took several actions to facilitate the development and implementation of the COMPASS system. In 2010, LDOE staff convened the Advisory Committee on Educator Evaluation (ACEE) to involve educators in the development of the system (LDOE, 2013b). In 2011, LDOE staff conducted several pilot tests of the three separate evaluative components (LDOE, 2013b). Then, the LDOE piloted the entire system during the 2011-2012 school year in “nine school districts and one charter school” (LDOE, 2013b, p. 4).

During the initial pilot year and the first six months of statewide implementation, LDOE staff trained all potential evaluators and solicited feedback from both evaluators and teachers (LDOE, 2013b). The feedback led to adjustments to the COMPASS rubric, including a reduction of the number of rating categories from 5 to 4. In response to feedback on the length of observations, the LDOE reduced the components of the original Danielson rubric to create the current COMPASS rubric (LDOE, 2013b). The COMPASS rubric contains 5 of the 23 performance items on the Danielson rubric (Danielson, 2011a, 2011b). However, unlike the STAR rubric (Teddle et al., 1990), the LDOE did not test the COMPASS rubric and its 5 performance items for inter-rater reliability or validity before it was implemented throughout the state.

LDOE staff tried to address concerns regarding reliability between evaluators by developing a video library for evaluators to connect their scoring decisions to examples of practice. LDOE staff also released each district’s scoring distributions as of late 2013 (LDOE,

2013b), however, LDOE staff did not release documentation regarding the frequency, content, or quality of this training given to principals by state or district officials.

The system included, as noted in Chapter 1, the development of a performance rubric for required annual classroom observations and the development of procedures to incorporate student performance data in determining teachers' final evaluation scores. The system included measures of student learning through the use of a value added measure (VAM) and student learning targets (SLTs). During the 2012-2013 school year, the LDOE calculated VAM scores for all math and English teachers in third through eighth grade and for all high school Algebra and Geometry teachers. Policymakers required teachers who were not eligible for VAM scores to set an SLT. LDOE statisticians calculated VAM scores with a formula that used annual student achievement on standardized tests to evaluate student growth (LDOE, 2013a). Beginning in the 2014-2015 school year, LDOE policymakers removed VAM from the COMPASS system, resulting in all teachers being evaluated by a classroom observation and an SLT.

SLTs are achievement goals which teachers set for their students. By considering details of their classes, such as the class average on a pre-test or the presence of students with exceptionalities, teachers set a specific level of student achievement that served as the student learning target for their end of the year evaluation. After giving an end of the year assessment, the principal rated teachers based on their classes' proximity to the initial SLT. The resulting score from either the VAM or the SLT ranged from 1.0 to 4.0 and was then averaged with the second component to create the overall rating (LDOE, 2013a, 2013b). Since SLTs were goals set by individual teachers and rated by principals on selected assessments, the goals, the assessments used to measure student progress, and the scores varied for each individual teacher.



The focus of this study was the qualitative component of the system, a performance evaluation in which principals used a new classroom observation instrument. This instrument, the COMPASS rubric, consisted of five performance items taken from the Danielson Framework along with four rating categories that ranged from Highly Effective to Ineffective. (Danielson, 2011a). Policymakers aligned the rating categories with a corresponding numerical range. For example, the range for the Ineffective and Highly Effective rating categories was from 1.0 to 1.5 and 3.5 to 4.0 respectively, a spread of one-half point. For the two middle rating categories, Effective Emerging and Effective Proficient, the range was from 1.5 to 2.5 and 2.5 to 3.5 respectively, a spread of one point each for these two rating categories. Although the COMPASS rubric contained less performance items for scoring than either the STAR or LCET rubrics, each performance item had a description for every rating category.

**COMPASS as a Standards-Based Rubric.** The resulting evaluation system of a classroom observation combined with a measure of student growth became known as COMPASS. LDOE policymakers described COMPASS:

COMPASS is the state’s educator support and evaluation system. The system is designed to provide all educators with regular, meaningful feedback on their performance and aligned supports to foster continuous improvement (LDOE, 2014).

Although the LDOE’s definition of COMPASS references both feedback and support, there was no clear policy or procedural description dedicated to feedback and support. Furthermore, the 2013 end of the year report on the COMPASS system noted “that there is more work to be done to ensure that teachers across the state were getting the feedback they need to drive gains in student learning” (LDOE, 2013b, p. 3).

The COMPASS rubric, unlike STAR or LCET, was a standards-based rubric that focused on describing and distinguishing levels of teacher performance (Danielson, 2011a; LDOE, n.d.).

A standards-based teacher evaluation rubric was one that Milanowski, Kimball, and White (2004) described as:

A comprehensive model or description of what teachers should know and be able to do, represented by explicit standards covering multiple domains and including multiple levels of performance defined by detailed behavioral rating scales. It typically requires more intensive collection of evidence, including frequent observations of classroom practice and use of artifacts such as lesson plans and samples of student work, in order to provide a richer picture of teacher performance (p. 2)

The Danielson Framework for Effective Teaching was one of the first standards-based rubrics to rise to prominence during the early 2000s (Heneman III et al., 2006). During the RTTT policy changes, many states adopted the Danielson rubric. In Louisiana, the rubric formed the basis for the COMPASS rubric for teacher evaluation (Danielson, 2011a, 2011b). School districts across the nation quickly adopted standards-based rubrics (Schmoker, 2012). Schmoker (2012) criticized the speed at which these rubrics were implemented after RTTT, where there was often a failure by states to pilot their new systems and instruments.

Standards-based teacher evaluation rubrics extended their scoring beyond the teacher and also analyzed the impact of the teacher on students' learning (Ingvarson & Rowe, 2008). For example, the rubric focused on instructional practices and required teachers to demonstrate student driven processes and procedures (Danielson, 2011a). Policymakers designed the COMPASS system as a standards-based teacher evaluation system, and its components and policies fit into the systematic criteria defined by Odden (2004):

A standards-based teacher evaluation system requires the following: 1. A set of teaching standards that describes in considerable detail what teachers need to know and be able to do. 2. A set of procedures for collecting multiple forms of data on teacher's performance for each of the standards. 3. A related set of scoring rubrics that provide guidance to assessors or evaluators on how to score the various pieces of data to various performance levels and a scheme to aggregate all microscores to an overall score for a teacher's instructional performance. 4. A way to use the performance evaluation results in a new knowledge- and skills-based salary schedule if the evaluation system is to be used to trigger fiscal incentives (p. 127).

LCET, by contrast, focused its evaluation on the teacher and included actions that were external to classroom activity, such as the teacher's dedication to professional development (LDOE, n.d.). Furthermore, policymakers did not develop a way to use the results of LCET to modify salary schedules.

### **Development of the COMPASS Teacher Rubric from the Danielson Framework.**

The foundation for the COMPASS teacher rubric is Danielson's 2011 edition of the *Framework for Effective Teaching* (LDOE, 2013b). Danielson's framework categorizes effective teaching into four domains comprised of twenty-two subcategories (Danielson, 2011a). These four domains were: planning and preparation, classroom environment, instruction, and professional responsibilities (Danielson, 2011b). During the evaluation, principals rated teachers on each subcategory across the domains using a scale ranging from 1.0 to 4 that corresponded with the following performance levels: unsatisfactory, basic, proficient, and distinguished (Danielson, 2011a).

Policymakers modified the original Danielson Framework to create an evaluation rubric that placed emphasis on student learning, as facilitated by teachers (Danielson, 2011a; LDOE, 2014). First, policymakers narrowed the focus of the Danielson Framework by moving from four domains and twenty-two subcategories to three domains and five subcategories (Danielson, 2011a). Then, policymakers eliminated the fourth domain of professional responsibilities and its subcategories, choosing instead to focus specifically on actions in the classroom (Danielson, 2011a).

In the first domain, planning and preparation, the COMPASS rubric only contained the subcategory of setting instructional outcomes (Danielson, 2011a). Policymakers omitted the following subcategories: demonstrating knowledge of content and pedagogy, demonstrating

knowledge of students, demonstrating knowledge of resources, designing coherent instruction, and designing student assessments (Danielson, 2011b). In the domain of the classroom environment, the sole focus was on the managing classroom procedure subcategory (Danielson, 2011a; LDOE, 2013b). Policymakers omitted the classroom environment subcategories of creating an environment of respect and rapport, establishing a culture for learning, managing student behavior, and organizing physical space (Danielson, 2011b). In the domain of instruction, the COMPASS rubric contained three Danielson subcategories: questioning and discussion techniques, engaging students in learning, and using assessment in instruction (Danielson, 2011a). In the domain of instruction, policymakers omitted two subcategories: communicating with students and demonstrating flexibility and responsiveness (Danielson, 2011b).

Louisiana policymakers also developed new practices and scoring methods for use with the COMPASS rubric, practices that aligned with the practices recommended by the *Measuring Effective Teaching* study (Cantrell & Kane, 2012). First, LDOE policymakers renamed the original categories from Danielson's work from unsatisfactory, basic, proficient, and distinguished to ineffective, effective emerging, effective proficient and highly effective (Danielson, 2011a; LDOE, 2013b). Second, policymakers assigned each category a numerical scoring value from 1.0 to 4.0. All subcategories were then averaged to determine the final annual evaluation score for a specific teacher (Danielson, 2011a). Furthermore, principals determined the final annual observation score from the average of multiple evaluations (LDOE, 2013b). In addition to the mandating of multiple observations, the LDOE required a pre-conference and post-conference meeting between principals and teachers.

## **Implementation of Teacher Evaluation Systems**

In this section, I provide a foundational understanding on the implementation of teacher evaluation systems. To understand the implementation of COMPASS within Louisiana school districts, literature on policy implementation at the school and district level must be examined. I examined the literature on policy implementation and then I analyzed the characteristics associated with professional development used to support policy implementation.

### **General Policy Implementation**

O'Toole (2000) defined policy implementation as “what develops between the establishment of an apparent intention on the part of government to do something, or to stop doing something, and the ultimate impact in the world of action” (p. 266). State officials started implementing COMPASS with policy mandates and training, before the new evaluation policies were put into practice by principals. State policymakers intended COMPASS to distinguish between levels of job performance, and the positive skew may have resulted from challenges during the implementation process.

Policymakers often struggle to control the factors related to successful policy implementation. McLaughlin (1987) described the challenges facing policymakers during the implementation process as “perhaps the overarching, obvious conclusion running through empirical research on policy implementation notes that it is incredibly hard to make something happen, most especially across layers of government and institutions” (p. 172). McLaughlin (1987) further explained, “at each point in the policy process, a policy is transformed as individuals interpret and respond to it” (p. 174), and also presented the implementation process as a continual process of bargaining and negotiating between policymakers, their policy mandates, and the individuals charged with implementation.

Policy implementation emerged as a field of public policy research during the late 1960s and early 1970s (Honig, 2006a; McLaughlin, 1987). Honig (2006a) defined the history of policy implementation research in terms of four phases: a focus on what gets implemented, attention to what gets implemented over time, growing concerns with what works, and confronting complexity. Honig (2006a) described the first phase as one in which researchers limited their examinations to the policy and its fidelity of implementation, then described the second phase as one in which researchers extended their studies to tracking implementation goals over an extended period of time. Regarding phase three, Honig (2006a) noted that investigators during the 1980s shifted their questions towards the factors that assisted or interfered with successful policy implementation. Finally, Honig (2006a) explained the current phase of research, wherein researchers recognized the complexity of the field. For example, researchers have moved beyond dichotomous debates and now acknowledge the strength and weaknesses of various approaches (Honig, 2006a).

In an analysis of district administrators, Honig (2006b) further examined the role of people in successfully implementing education policy, recognizing boundary spanners, individuals who work directly on policy implementation among stakeholders, as an essential component of successful policy implementation. Honig (2006b) stressed the need for administrators with varied experiences to serve as agents of change and implement policy changes within school districts. However, Honig (2006b) found that those who worked directly to implement policy encountered difficulties balancing their dual responsibility of serving within the district office and supporting school sites.

In a paper presented to the American Education Research Association, Darling-Hammond (1992) advocated for a paradigm shift in policy implementation within schools.

Specifically, Darling-Hammond (1992) asked policymakers to focus on building the capacity of individuals to ensure sustained reforms in schools. Fullan and Quinn (2010) defined capacity as the “capability of the individual or organization to make the changes required and involves the development of knowledge, skills, and commitments” (p. 1). Darling-Hammond (1992) noted the importance of capacity building when she described it as “required for a reform that must rely on the transformative power of individuals in schools to rethink their practice and redesign their institutions” (p. 7).

The concept of capacity assisted my understanding of both the factors influencing principals’ rating decisions as well as the support they received during implementation. Generally speaking, there may have been factors within rural school districts that diminished the capacity of principals’ accurately rating teachers on the COMPASS rubric. Furthermore, during the implementation of COMPASS, both the state and district may have sought means to provide capacity, for example, through professional development and policy support.

Work by Hord, Rutherford, Huling-Austin, and Hall (1987) also framed change within schools as a process that depended on individuals; they studied individuals, known as change facilitators, who supported organizational changes. Change facilitators engaged in actions such as training others and monitoring changes; in their studies of change facilitators in schools, Hord et al. (1987) noted the principal as the primary change facilitator, often with secondary supporters from their staff or the district.

From an examination of the literature, researchers in the field of policy implementation have focused on the individuals who were involved in the implementation process. When individuals had the capacity to support policies, they could ensure proper implementation through steps such as training others. With COMPASS, policymakers made principals

responsible for evaluating teachers. If principals did not have the capacity to evaluate teachers, then principals could not have evaluated teachers in accordance with the standards of state policymakers. Researchers have noted the need for high-quality professional development to support new systems of teacher evaluation and to ensure accurate evaluation of teachers by principals (Cantrell & Kane, 2012).

### **Professional Development to Support Implementation**

Before the current push for teacher evaluation, researchers examined high-quality professional development for educators (T. Guskey & Sparks, 2002; T. R. Guskey, 2002). By utilizing the general educational literature on professional development, I identify specific general guidelines for high-quality professional development and compare these features with the support related to COMPASS. If the support offered to implement COMPASS was inadequate, then the LDOE policymakers would not realize their scoring expectations on a systematic level.

**Characteristics of High-quality Professional Development.** In a survey of teachers concerning their professional development experiences (Garet, Porter, Desimone, Birman, & Yoon, 2001), investigators identified factors such as duration, a focus on content, active learning, and coherence with daily responsibilities as influences on altering teacher practices. Other researchers placed an emphasis on content and duration, and also promoted the use of collaborative learning experiences (Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009). Approaches such as the consensus model (Elmore, 2008) incorporated both collaboration and adherence to recognized adult learning theories into aspects of effective professional development for teachers.



The consensus model offered a broad view of high-quality professional development by bringing together several concepts from years of literature and research. This resulted in a list of factors associated with high-quality professional development (Elmore, 2008). From this examination of literature on professional development, three overarching aspects emerged: duration, collaboration, and cohesion of content (Darling-Hammond et al., 2009; Garet et al., 2001; T. Guskey & Sparks, 2002). A research team led by Darling-Hammond noted these four aspects in a comprehensive examination of effective professional development practices:

Effective professional development is intensive, ongoing, and connected to practice; focuses on the teaching and learning of specific academic content; is connected to other school initiatives; and builds strong working relationships among teachers (Darling-Hammond et al., 2009, p. 5).

If principals did not receive high-quality professional development on rating classroom practice using the new COMPASS rubric, then they may be unable to accurately utilize the instrument (Ho & Kane, 2013).

***Duration.*** Duration of professional development related to two seemingly independent focuses: the length of support for professional development practices and the number of hours spent in direct training (Garet et al., 2001). Garet et al. (2001) extolled duration as beneficial in providing greater opportunities to influence the transfer of training. Garet et al. (2001) specifically noted the duration of professional development as a key determinant of quality:

The fact that both time-span and contact hours have independent effects on our measures of core features suggests that both dimensions of duration were important. Professional development is likely to be of higher quality if it is both sustained over time and involves a substantial number of hours (p. 933).

The duration requirement, especially measured by contact hours, is a facet of high-quality professional development that educators seldom experience (Darling-Hammond et al., 2009).

The number of contact hours required for effective duration may be as high as fifty hours a year

per content focus (Darling-Hammond et al., 2009). In a comprehensive analysis of the literature on professional development (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007), this minimal number of hours amounted to at least fourteen contact hours. The length of professional development experiences was also noted by Yoon et al. (2007) as having a positive impact on the effectiveness of training. Even if the required contact hours for effective professional development varied in the literature, it was clear that most educators did not have access to enough hours of professional development (Darling-Hammond et al., 2009).

Ho and Kane (2013) demonstrated a need for professional development in assisting principals to make accurate scoring decisions. In Louisiana, districts approved their principals as evaluators after training, however, there was no requirement for a specific number of hours of professional development for principals. It could not be assumed that, because principals previously conducted observations, they would automatically understand how to use a new instrument. Furthermore, principals in Louisiana had to make adjustments from the local scoring norms of LCET to the statewide expectations of COMPASS.

***Collaboration.*** Research (Hunter & Back, 2011; James-Ward, 2011; Joyce & Showers, 2003) identified peers as individuals who can supply additional support in the form of mentoring, coaching, or peer learning. If well-defined systems of school level collaboration were in place, then educators who received training may have been more able to effectively train their peers (Garet et al., 2001). Darling-Hammond et al. (2009) also recommended collaboration, noting that “collaborative approaches to professional learning can promote school change that extends beyond individual classrooms” (p. 5).

The purpose of collaboration was to increase the knowledge and skills of principals by learning from other administrators (Ashton & Duncan, 2012; Brody, Vissa, & Weathers, 2010).

Peer collaboration for principals can be a difficult challenge, especially in rural districts that contain few administrators over a large geographical area (Ashton & Duncan, 2012; Saunders, Goldenberg, & Gallimore, 2009). Supovitz, Sirinides, and May (2010) noted that the importance of collaboration must be an emphasis at all levels of education. Professional development for principals should have fostered collaborative practices between administrators at different schools (Supovitz, Sirinides, & May, 2010). In most school districts, collaboration for principals took place in mentorships between experienced and inexperienced principals (T.L. Alsbury & D.G. Hackmann, 2006; Ashton & Duncan, 2012).

Regarding ensuring the validity and reliability of ratings, Ho and Kane (2013) emphasized the need for principals experienced in scoring teacher performance to share their scoring decisions with other principals. Both the reliability and accuracy of scoring decisions increased when these decisions were made by teams of principals who discussed and justified their rating decisions (Ho & Kane, 2013). There was no indication that COMPASS training utilized peer training, such as mentorship from an experienced evaluator, on the scoring expectations of the COMPASS system.

***Cohesion of Content.*** Joyce and Showers (2003) emphasized a focus on content and recommended that the purpose and knowledge taught must be aligned to address issues of practice. Furthermore, Garet et al. (2001) promoted cohesion of content as a response to the lack of cohesion between individual professional development events such as workshops and the daily practices of educators. In order to train principals in accurate and reliable scoring, training sessions on the instrument must have clearly communicated examples of classroom performance and the expected scores associated with each example (Cantrell & Kane, 2012; Ho & Kane, 2013). One recommendation by Ho and Kane (2013) was that evaluators practice their scoring

decisions with the instrument on videos of classroom instruction until they were able to accurately and reliably evaluate teachers. If principals were not provided relevant opportunities for testing their scoring decisions in their professional development on COMPASS, then they may have needed training on accurate and reliable rating. The historic practices of principal professional development should be analyzed with the three aspects of high-quality professional development established from the literature.

**Principal Professional Development.** After an analysis of both the literature on principal professional development and on training principals to accurately rate classroom instruction, two distinct methods emerged: mentorships and training sessions. In the context of training principals on the use of a new observation instrument, mentorships between experienced and novice evaluators were studied as a means of increasing reliability between raters (Cantrell & Kane, 2012; Ho & Kane, 2013). Ho and Kane (2013) noted the use of training sessions featuring videos of classroom practice, with practiced rating by participants and discussion of scoring decisions, as a method of producing a consensus in scoring decisions amongst principal.

In terms of this analysis, policymakers should have examined each practice for its prevalence of use, its presumed effectiveness, and its tendency towards alignment with the three aspects of duration, collaboration, and cohesion. Each of these practices were avenues by which principals were traditionally exposed to new information on a professional level (Ashton & Duncan, 2012).

**Mentorship.** Mentoring and coaching had a history of being recognized as effective professional development programs and as means to guide and assure the implementation of new skills and practices into an individual's professional practice (Joyce & Showers, 2003). Mentorships should not be confused with internships. For principals, internships were mostly a

pre-service training method used in leader preparation programs, while mentorships were fostered in a structured manner after an individual had been hired as a principal (Alsbury & Hackmann, 2006). Mentorships served a prominent role in principal professional development. As Alsbury and Hackmann (2006) noted: “the use of mentoring relationships to facilitate and sustain professional development is an age-old tradition” (p. 1). For example, in 1997, the state of Kentucky mandated that all new principals were to complete a mentorship under a more experienced principal, with the entire process monitored and guided by a committee of educators (Wells, Rinehart, & Scollay, 1999).

Mentorship programs were often sustained throughout the first few years of a principal’s practice (Alsbury & Hackmann, 2006; Wells et al., 1999). This sustained time period should allow for a greater transfer of skills and knowledge from the mentor to the mentee. Mentorships were also inherently collaborative, specifically in the relationship between the participant and their more experienced mentor. In addition, some mentorship programs incorporated collaboration among multiple principals who receive training from a district supervisor (Alsbury & Hackmann, 2006). Survey responses from mentorship participants revealed a positive association of mentorship with both impacting student learning and increasing collaboration amongst principals (Howley, Crowley, & Howley, 2002). Finally, when mentorships are especially structured by a district or state agency, they also benefit from an alignment with the immediate needs of novice administrators as well as current standards for achievement and accountability.

Ho and Kane (2013) recognized mentorship from an experienced classroom evaluator as a means of increasing the reliability of principals in their scoring decisions. A mentor allowed a principal to adjust their own rating decisions to match that of an experienced rater, which

increases the ability of the principal to accurately assess classroom instruction (Cantrell & Kane, 2012; Ho & Kane, 2013). Ho and Kane (2013) indicated this direct relationship between experienced and novice evaluators was an essential step in ensuring that new observational instruments were adopted properly. During COMPASS implementation, policymakers did not require the site-based mentorship as a training method. Instead, COMPASS training consisted of individual workshops and informational meetings.

***Training Sessions.*** Results from a survey of principal professional development experiences showed that the practice of an intensive training series was not a common practice of districts or states (Elsberry & Bishop, 1993). Only 31% of the responding principals participated in a program of any meaningful duration before the start of the school year; instead, a vast majority of those same principals attended single session workshops on various topics throughout the year (Elsberry & Bishop, 1993). These single session workshops were not thematically connected, nor reported as effective training (Elsberry & Bishop, 1993). Training sessions experienced by principals often failed to meet the criteria of duration and cohesion.

In light of the research on duration, cohesion, and collaboration, some district leaders began systematic changes in the delivery of training sessions for principals. Grodzki (2011) reported positive results from summer training series for principals that were of significant duration and were collaborative in nature. From a survey on effective district training sessions, Grodzki (2011) noted that “candidates reported that they found these sessions rewarding and enjoyed the opportunity to connect with other candidates and meet district staff” (p. 10). Nomore (2007) described effective programs for principals as ones that were of exceptional duration, particularly those that continue throughout the year.

However, the majority of training sessions may not constitute high-quality professional development experiences. For example, in a survey of first year principals, Howley, Chadwick, and Howley (2002) found that only half of first year principals reported that they learned valuable information in workshops, though nearly all experienced single session workshops in their professional development experiences. This information coincided with historically reported data from Elsberry and Bishop (1993), who found that although 92% of principals experienced single session workshops, only 8% of principals designated these workshops as their choice for the most effective training practice. Regarding training sessions for principals, effective training sessions were less prevalent than single session workshops that lack any of the features of high-quality or effective professional development.

### **Professional Development in Support of COMPASS**

As with the implementation of any new policy or procedure, a new system for evaluating teachers requires adequate training and support to achieve fidelity of implementation. The importance of proper training for principals was one of the primary recommendations of a collection of studies funded by the Bill and Melinda Gates Foundation. These studies advocated for substantial, high-quality training on observational practices to improve rater reliability (Cantrell & Kane, 2012). Principals should be trained to a point at which they were consistently able to demonstrate competency in distinguishing between levels of proficiency in teaching practice (Cantrell & Kane, 2012).

Louisiana school districts varied their support and training sessions for principals. Policymakers did not require support and training for COMPASS when the legislature altered its policies on teacher evaluation (Act No. 54, 2010). During the summer of 2012, LDOE staff held regional meetings for district supervisors to train district staff on the new instrument, with the

understanding that these district administrators were responsible for training and certifying school administrators as competent evaluators (COMPASS, 2013). After the training from district administrators, principals gained certification and rated teachers and submitted rating data into the state's online reporting system (COMPASS, 2013; LDOE, 2013b).

Training sessions, specifically those in which multiple principals collaboratively rate videos of classroom instruction and receive direct feedback on their scores, can increase both accuracy and reliability of scoring decisions (Ho & Kane, 2013). However, Ho and Kane (2013) also recognized that many principals do not have access to the level of training required to improve their scoring decisions: "it would be helpful to provide better training and certification tests for prospective raters, followed by regular efforts to ensure that observers remain calibrated to the standards" (p. 30). By primarily relying on training sessions to support COMPASS implementation, LDOE staff may have not exposed principals to high-quality professional development related to the COMPASS rubric.

In the absence of any statewide, unified training protocol, the LDOE allowed school districts to determine the extent and nature of any training for principals (COMPASS, 2013). Without unified training, there may have been issues with establishing the reliability of ratings among school districts, as each school district would have varying degrees of training and expectations of scoring norms.

### **Conceptual Framework: Culture and Context**

I examine organizational culture and context in the section, and provide a conceptual framework for understanding the factors that may influence principals. I also detail specific challenges faced by rural schools and their personnel, including population decline, financial issues, access to resources, and difficulty in attracting and retaining teachers.



## **Organizational Culture**

In COMPASS, teachers were rated by the principals of their school, and not by a team of external evaluators. As such, the culture and context of the school or local district may influence the use of COMPASS, as principals and teachers shared a cultural space within the school. Given the many definitions and perspectives on culture, culture was often contextually defined in each area of research (Honold, 2000). Louis (2006) recognized schools as organizations with their own distinct cultures. Furthermore, the rural context created cultures unique to rural schools (Theobald & Nachtigal, 1995).

The concept of culture in organizations developed as organizational researchers began to apply anthropological concepts to their theories on the behaviors of individuals in organizations (Smircich, 1983). For example, anthropological knowledge of the use of symbols as cultural artifacts could provide perspective on the symbols inherent in organizations (Smircich, 1983). A common view of culture in organizations was that culture was both a product of the organization's structure and an influential variable on the thoughts, behavior patterns, and actions of those in the structure. As Smircich (1983) explained:

Organizations are seen as social instruments that produce goods and services, and, as a by-product, they also produce distinctive cultural artifacts such as rituals, legends, and ceremonies (p. 344).

These organizational products produced “shared key values and beliefs” (Smircich, 1983, p. 345), manifesting a common culture among the individuals in the organization. In relation to this study, the structure and environment of rural schools created a unique culture. In turn, this culture influenced the behavior of all members of the organization; policy implementation in organizations is carried out by individuals who are influenced by the institution's culture and the environmental context (Hatch, 2012). Furthermore, the artifacts produced by the functions of the

organization, such as ratings of classroom performance, were assimilated into the organizational culture and influence future decisions. While many features of an organization, such as rituals, traditions, processes, artifacts, and statements of belief, were aspects and evidence of culture, the aspect of culture of most interest to this study was that of behavioral norms. Norms are specific patterns of behavior enforced by the culture of an organization (Knutson et al., 2010). For example, the cultural expectation of how principals should score teachers during their classroom observations was a behavioral norm. Accordingly, the potential of culture to influence the rating decisions of principals warranted an examination of theories of culture in organizations.

In their work on theories of culture in organizations, Denison and Spreitzer (1991) divided organizational culture into four orientations: group, developmental, rational, and hierarchical. Each orientation placed value on a different aspect of the organization. For example, the group orientation valued the members of the organization, whereas the developmental orientation valued competition with external sources, the rational orientation valued performance, and the hierarchical orientation valued the structure and stability of the organization (Denison & Spreitzer, 1991). Denison and Spreitzer (1991) theorized that each of these types of cultures were both present in an organization and in conflict with each other:

Each of the cultural orientations has a polar opposite. The group culture, which emphasizes flexibility and an internal focus, can be contrasted with the rational culture, which emphasizes control and external focus. The developmental culture, which is characterized by flexibility and an external focus, can be contrasted with the hierarchical culture, which stresses control and an internal focus (p. 6).

From this theory of cultural conflict and competing orientations, Denison and Spreitzer (1991) described culture in an organization as multifaceted. In an organization, there may be competing orientations of culture which drive action and craft responses. For example, an organization with a strong group or hierarchical cultural orientation could resist an external policy initiative, while

an organization with a strong rational or development orientation could be more accepting of the policy. In relation to the implementation of COMPASS in rural schools, rural schools could tend towards internal orientations of culture, which could have lead principals to protect their teachers from consequences related to poor performance on the COMPASS rubric.

Researcher and theorist Karl Weick provided two additional perspectives on culture and its influence within organizations in his work on loose coupling (Orton & Weick, 1990) and sensemaking (Weick, 1995). The term coupling was used to explain the operational, communicative, or cultural separation between groups in an organization or between different organizations. Orton and Weick (1990) used the term loose coupling to explain a break in the cultural identity between two entities. In his early work, Weick (1976) described organizations in larger educational structure as being loosely coupled. For example, a rural school and an urban school in the same school district could be loosely coupled, as, although they were schools in the same district, their cultures were distinct. Further explanation by Weick (1976) described these cultural distances as creating opportunities for “dysfunctions associated with loose coupling” (p. 18). Relative to COMPASS, the expectations and culture of the LDOE in regards to scoring expectations for principals may have been loosely coupled to the expectations at the district level or in rural schools. Loose coupling between levels of organizations may create gaps in communication and cultural expectations that may have led to dysfunction in how COMPASS was implemented.

Weick’s views on culture also extended to the purpose and activity of culture in organizations. Weick (1995) described a process known as sensemaking, which was a continual process both contributing to and influenced by the culture of an organization. As Weick (1995) explained, sensemaking allowed individuals to process their activities and roles in an

organization in a manner that was “socially acceptable” (p. 61). Weick (1995) linked aspects of culture, such as myths and rituals, to attempts at sensemaking by the organization and its members. In relation to the present study, principals went through a sensemaking process to reflect, understand, and accept their scoring decisions, and cultural expectations influence this sensemaking process.

Swidler’s (1986) cultural theory of settled versus unsettled interactions was based on the influence of culture during certain phases of an organization’s lifespan. During settled phases, cultural influences were diminished and difficult to isolate (Swidler, 1996). The culture of a settled and established organization allowed individuals to act in a manner that was more independent from the norms of the organization. During times of uncertainty, individuals in organizations tended to avoid individualistic actions and engage in actions aligned with cultural norms (Swidler, 1996). An example of an unsettled organization was one that faced an external threat or was undergoing a change. The implementation of COMPASS was a change from LCET and a return to systematic teacher evaluation last seen with LATEP and STAR. The act of implementing COMPASS was an unsettling act for a school. During this unsettled period, individuals would be more inclined to follow the culture of an organization, rather than their own individual decisions. Furthermore, the cultural norms in Louisiana schools were established norms in relation to teacher evaluation and the scoring of classroom observations from the long-term use of LCET. Principals may have been inclined to rely on these previous cultural norms when faced with uncertainty in scoring teachers on the COMPASS rubric.

To understand how culture influences the behavior of individuals, it was necessary to explore a framework that examined organizational culture at multiple levels (Schein, 2004). Schein (2004) provided a framework with three levels of culture: artifacts, espoused beliefs and

values, and underlying assumptions. Artifacts are highly visible aspects of culture and may include structures and processes in an organization (Schein, 2004). For example, an organization's constitution, by-laws, or charter, are defined as artifacts (Schein, 2004). Espoused beliefs and values are aspects of an organization's culture that members of the organization voice that are also subject to conflict from internal and external sources (Schein, 2004). When members find that a method, action, or belief system has utility, solves conflicts, or demonstrates some advantage to the organization, it becomes an espoused belief (Schein, 2004). Finally, Schein (2004) indicated that individuals often do not see the underlying assumptions of the group, and, therefore, these assumptions are difficult to confront or change. The difficulty in addressing underlying assumptions comes from their acceptance as core, foundational values which grants them immunity from debate, discussion, or reflection by members of an organization (Schein, 2004).

All three levels of culture have a direct, reciprocal influence on each other, according to Schein (2004). For example, cultural artifacts influenced the creation of publicly shared values that individuals ultimately internalize. These publicly shared values contribute to a set of underlying assumptions. Individuals' underlying assumptions influence the beliefs that they are likely to espouse to others. The organization subsequently adopts these beliefs, which become visible through artifacts the organization produces.

For this study, Schein's (2004) levels of culture enabled the examination of the behaviors of principals in their evaluation of teachers. The three levels of culture simultaneously influence the behaviors of principals, as principals were members of the organizational culture of a school. Conflict between the levels of culture and the policies of COMPASS may have influenced the behavior of the principals in their rating decisions during classroom observations. For example,

principals' underlying assumptions regarding effective teaching or the COMPASS rubric may have influenced their scoring of teachers. These underlying assumptions may have been in conflict with the espoused beliefs of the school leaders regarding the effectiveness of their teachers, the performance of their school, or the COMPASS system. Furthermore, due to the loose coupling found between various organizational levels, the norms from the local district, in terms of their previous evaluation policies and procedures, may have been different from the LDOE's norms and beliefs regarding COMPASS.

From the literature on culture, principals are both influenced by and influential on the local culture during the process of teacher evaluation. The implementation efforts of the LDOE may have been insufficient for creating a new culture of scoring expectations for numerous reasons. First, state policymakers may have faced difficulty in communicating the expectations due to the loose coupling between the local districts and the LDOE. Second, the complex structure of culture may have made it difficult to transition between espoused beliefs of the LDOE to the underlying assumptions regarding teacher effectiveness found in rural schools. Third, the accepted scoring norms associated with LCET may have required sustained, high-quality professional development to address and assimilate new norms associated with COMPASS. An exploration of culture provided possible explanations for the positive skew found in the COMPASS results and also provided a framework for understanding how principals explained their scoring decisions.

### **Organizational Context**

In addition to culture, the context in which organizations operate was an important aspect that warrants examination. Context is much more than the physical environment of the organization, as the concept of context in organizational literature encompassed both the

environment and how the organization was situated in that environment (White, 1986). The strategy, structure, or leader of an organization (White, 1986) influenced the context in which the organization operates. Ghoshal and Bartlett (1994) recognized that context influences all the members of an organization, including the leaders of an organization (Benson, Saraph, & Schroeder, 1991). School principals may have been more aware of and influenced by the context in which their schools operate, as they must deal with both internal and external factors.

Context had both internal and external dimensions in relation to organizations. Culture was an aspect of the internal organizational context (Bloor & Dawson, 1994). Both culture and context were recognized (Bloor & Dawson, 1994) as having mutual influence on each other. While culture was a social structure or belief system shared by individuals in an organization, context was the holistic social, physical, and policy environment in which the organization operates. S. Kim and Lee (2006) also identified the structure, knowledge, and culture of an organization as aspects of the internal context of an organization. Therefore, every organization and its members are influenced by the context in which they operate.

In relation to COMPASS, a principal must have considered the policies of their school district, the culture of their school, and the environmental context of the school. Since context was multifaceted, conflicts may arise. For example, if a principal did not receive adequate support to assist ineffective teachers, then that principal may have defied scoring norms and avoided giving an Ineffective rating to a teacher. In summary, the context in which a principal operated in their district and local environment may have influenced their decisions.

Of interest to this study was the context faced by principals of rural schools. From the COMPASS results, rural principals were less likely to label teachers as ineffective. There may have been factors unique to the context of rural districts that influenced principals and their

scoring decisions on the COMPASS rubric. One possible factor may have been the conflict between rural school district personnel needs and the policies regarding ineffective teachers. The policies applicable to ineffective teachers in the COMPASS program designated specific courses of action by the district personnel that began with mandatory retraining of teachers known as intensive assistance (Act No. 54, 2010; LDOE, 2013b). If the ineffective rating continued for a teacher in subsequent observations, the initial policy mandated termination (Act No. 54, 2010). Given the difficulty in recruiting teachers to rural schools, principals in rural districts may have been reluctant to identify teachers as ineffective (Morton & Harmon, 2011). Principals in rural districts also faced a lack of administrative support due to the large geographical area of the school district combined with a smaller district support staff (Ashton & Duncan, 2012). Principals may not have received sufficient professional development, training, or directives from their district administrators during the implementation of new programs and policies (Ashton & Duncan, 2012). The rural context and its factors related to the COMPASS system are explained further in the proceeding section.

### **Factors that Influence Rural Principals**

The preceding sections analyzed factors that were unique to rural schools. As part of the broader context in which rural principals operate, these factors may influence the implementation of the COMPASS system or the scoring decisions of principals. These contextual factors also have a reciprocal relationship with rural culture, in that they were both influential to the local culture and influenced by the local culture. However, to explore the factors that influence rural principals, the concept of rural must be defined in specific terms.

**Definition of Rural.** Competing physical, cultural, and philosophical criteria by which districts could be categorized make it difficult to define a district as rural (Stelmach, 2011). For



example, Stelmach (2011) gave credence to socially constructed, cultural definitions of the concept of rural. Morton and Harmon (2011) associated the concept of rural with both geographical and cultural isolation. Other definitions focused on the sociological and cultural cohesion and connectedness found in small towns (Wake, 2012). A researcher could also categorize rural districts by physical and functional determinants, such as the population density of an area (Stelmach, 2011). Other researchers define rural areas through the distance between an area and the nearest commercial center or city (Morton & Harmon, 2011).

The standard definition regarding rural schools and students in the field of education came from a report by the National Center for Education Statistics (2007). This report defined schools as “rural” based on their distance with rural areas exceeding a distance of five miles from the nearest commercial center (Provasnik et al., 2007; Wake, 2012). From this categorization, researchers established the prevalence of rural schools to be around 30% of all American public schools (Ashton & Duncan, 2012; Provasnik et al., 2007; Wake, 2012).

The term “rural” is not a homogenous characterization. In the most rural areas, school districts contain less than a thousand students spread through geographical areas with population densities of less than five people per square mile (Morton & Harmon, 2011). Researchers defined these areas as remote or frontier, in that there was a distance greater than 25 miles between the area and its nearest commercial center (Provasnik et al., 2007; Wake, 2012). For the purpose of this study, the term “rural” was based upon population density and distance to a metropolitan area, in accordance with the National Center for Educational Statistics definition (Provasnik et al., 2007).

**Context of Rural School Systems.** While there may have been some differences in the physical characteristics of areas that fall under the established definition of rural by the National

Center for Educational Statistics, literature in the field (Ashton & Duncan, 2012; Maranto & Shuls, 2012; Provasnik et al., 2007) described several contextual challenges common to most rural districts. These issues, such as population decline, financial difficulties, access to resources, and the ability to recruit and retain effective teachers, were challenges which schools and districts faced in rural areas (Morton & Harmon, 2011; Nichols, 2004; Provasnik et al., 2007; Stelmach, 2011). These issues challenge the stability of rural communities (Provasnik et al., 2007). Furthermore, researchers described these issues as increasing in their severity over time (Stelmach, 2011). All of these factors impact the experiences of educators and students in rural school districts.

**Population Decline.** By their nature, rural areas have a low population density (Provasnik et al., 2007), however, the social and physical environments of a rural area are more complex than the size of the population. Rural areas are less densely populated areas, and many also experience a continual decline in population (Stelmach, 2011). In a recent survey, rural teachers cited population decline as their primary concern (Morton & Harmon, 2011). The decline in student population has forced districts to consolidation and close schools. These decisions destabilized the rural school system (Wake, 2012). School closure and consolidation create a smaller number of schools serving a more disparate population of students in a larger geographical area, which further contributed to community isolation and impeded efforts to build a sense of community (Morton & Harmon, 2011; Wake, 2012).

With a general trend of a population exodus in rural areas, rural school districts faced continual staff turnover (Maranto & Shuls, 2012). With a limited pool of candidates, principals may have been reluctant to rate a teacher as Ineffective and to begin the mandated process to terminate ineffective teachers. Instead, principals may have felt pressure to try to maintain

existing staff and focus on their improvement rather than their replacement. In some cases, given a lack of candidates, replacing an ineffective teacher may not have been an option.

**Financial Difficulties.** In a survey of teachers in geographically isolated, rural Montana schools, nearly half of the respondents cited financial difficulties faced by the rural school system as their greatest concern (Morton & Harmon, 2011). The absence of sufficient school funding was often representative of the general economic decline, poverty of the population, and the lack of tax base present in rural communities (Voyles, 2012). The standard reliance of schools on local funding sources placed rural schools in a difficult financial situation when compared to schools in other settings. According to the National Center for Educational Statistics (2007):

Rural public schools depend more on state funding than city and suburban schools (which tend to receive a greater proportion of their funding from local sources) and...rural public schools spend more per student than public schools in cities, suburbs, and towns when adjusted for geographic cost differences (p. 79).

Issues of school funding were also tied to declining school enrollment, as student enrollment was one of the primary determinants of state funding for school districts (Morton & Harmon, 2011). Rural school systems may be stuck in a complex cycle of financial decline, where declining enrollment and a lack of funding for school systems are both a cause and a result of continual systematic decline and instability (Stelmach, 2011).

Financial challenges increased the severity of other issues as well. With rural districts facing financial difficulties, principals may have been unable to obtain training and professional development for themselves (Ashton & Duncan, 2012; Howley et al., 2002). Rural principals were also generally less experienced and less educated than non-rural principals (Howley et al., 2002). Rural principals often need additional professional development to perform satisfactorily in their positions (Howley et al., 2002). Financial challenges in rural school districts also result in

a greater difficulty in incentivizing the recruiting of qualified teachers (Maranto & Shuls, 2012). Without necessary financial support, Louisiana principals in rural districts may have been unable to obtain access to the additional experience, education, and training needed to properly evaluate teachers. In addition, the financial situation of rural districts may present challenges in attracting effective teachers. Principals may have had to consider their district's inability to financially incentivize the recruitment of teachers when rating their current staff.

**Access to Resources.** The trending population decline and the distance between communities contributes to a prevailing sense of personal, professional, and cultural isolation (Ashton & Duncan, 2012). Isolation takes many forms and has many impacts. For example, the great distances between families and the schools their children attend can result in a lack of parental and community involvement at the rural school site (Voyles, 2012). Physical distance could also create high rates of absenteeism and low graduation rates in the student population, as transportation to a school site and a sense of cohesiveness in the school community may be difficult to guarantee (Provasnik et al., 2007; Stelmach, 2011; Voyles, 2012). This sense of isolation was cited by researchers as a significant challenge for those who were unable to take advantage of opportunities for professional collaboration, mentorship, and learning (Ashton & Duncan, 2012; Morton & Harmon, 2011).

According to Cantrell and Kane (2012), professional development is necessary to train principals to accurately and reliably rate the job performance of their teachers. Even in the absence of financial difficulties, the geographical and social isolation of a school district creates an environment that is not conducive to collaboration between educators at all levels of the system (Ashton & Duncan, 2012). Without collaborative training exercises or mentorship from

experienced evaluators, rural principals may have been unable to make accurate scoring decisions (Ashton & Duncan, 2012; Cantrell & Kane, 2012; Howley et al., 2002).

**Difficulty Attracting and Retaining Teachers.** The trend of declining population also influences a situation reported by many rural school systems: the difficulty of attracting and retaining teachers (Morton & Harmon, 2011). The economic opportunities present outside of rural areas exacerbate the population decline, especially among educated members of the rural community (Wake, 2012). Compared to urban and suburban school districts, teacher shortages were much more likely to occur in rural districts which feature teacher salaries below their state's average (Maranto & Shuls, 2012). These teacher shortages were most pronounced in subjects requiring certification in science, foreign language, and math (Goodpaster, Adedokun, & Weaver, 2012; Maranto & Shuls, 2012). However, even if financial compensation was not an issue, teachers may have been reluctant to move to rural environments due to the geographical, social, and cultural isolation faced by educators in rural districts (Maranto & Shuls, 2012). Thus, many rural school districts faced chronic teacher shortages from their inability to attract and retain teachers (Stelmach, 2011).

In the face of a teacher shortage, rural educational leaders often had to make less than optimal hiring and retention choices. As Maranto and Shuls (2012) wrote:

Principals worry about simply filling vacancies, not hiring the best teachers. The inability of some rural and urban schools to attract applicants leaves principals in the precarious position of having to hire whoever walks through the door, or failing to offer some courses. (p. 1)

In turn, the inability to hire effective, qualified educators prevented schools from offering quality instruction and course options as teachers were needed to teach courses in which they had little training, experience, or interest (Goodpaster et al., 2012). Furthermore, struggling educators

may have been unable to access fundamental resources and professional development given the geographical isolation of many rural areas (Goodpaster et al., 2012; Howley et al., 2002).

Researchers also focused on the retention of teachers in rural school districts (Maranto & Shuls, 2012; Provasnik et al., 2007). Martano and Shuls (2012) described the causes of teacher turnover in rural districts in terms of the demographics of the rural populations:

Not surprisingly, turnover of teachers was connected to the demographics of the students they teach, including achievement level. Teachers often leave low-paying, low-achieving schools in favor of employment in high-paying, high achieving schools. (p. 2)

Other research by Goodpaster et al. (2012) confirmed this finding with surveys of rural educators citing low pay and poor student performance as the most significant negative factors that they had to deal with during their tenure as rural teachers. However, the difficulty of teacher recruitment and retention may have gone beyond the financial difficulties faced by a rural school district. Faced with existing difficulties in recruiting and retaining teachers, principals may have made scoring decisions that focused on maintaining existing staff, rather than accurately addressing issues in job performance.

Overall Environment. The contextual factors faced by rural school systems did not act in isolation, but rather were interconnected. As the population declined, revenue fell, and as the district's revenue fell, so did the rural district's ability to compete for teachers with higher paying districts. The combination of these environmental conditions gave an image of rural school systems as facing increasing threats to their ability to effectively educate students. While some educators found positive aspects of rural culture, such as the social cohesion and shared cultural experience, the general consensus was that rural districts faced serious difficulties in maintaining a qualified teaching staff (Maranto & Shuls, 2012; Wake, 2012). Rural principals must deal with these challenging conditions and seek stability for their schools (Voyles, 2012).

The implementation of policy initiatives in rural schools did not occur in isolation. Rural issues may influence policy implementation to such a degree that the intent of the policy may fall short of its outcomes. COMPASS had clear consequences for teachers based on the rating decisions of principals, and may have faced additional challenges in how it was implemented by rural district principals, whose scoring decisions influenced the stability of their staff.

### **Summary**

From the literature, the field of teacher evaluation has evolved to evaluating job performance based on standards of performance. Since RTTT, teacher evaluation has been a central point of national policy, and state policymakers developed and implemented their own systems. In Louisiana, COMPASS was developed to evaluate teacher job performance in adherence to the selection criteria of RTTT. However, the positively skewed results of COMPASS warrant investigation, since the results were consistent with previous, non-standards-based systems. These positively skewed results were most notable in rural school districts, where principals throughout the districts did not rate any teachers as ineffective. With such large-scale patterns, factors may have influenced the rating decisions of principals outside of the considerations of the COMPASS rubric.

An investigation of the possible factors which influenced the rating decisions of principals began with a conceptual framework of culture and context in an organization, such as a school. Culture may have influenced principals on many different levels, such as public artifacts and hidden belief systems. Previous rating practices may have exerted a hidden influence which conflicted with the LDOE's expectations regarding scoring norms. Furthermore, the context in which rural schools operated could have been an influential factor. For example, the difficulty in attracting teacher candidates to a rural district may have influenced principals to

avoid rating teachers as Ineffective, which would have resulted in sanctions against their position.

Although possible factors have been examined, this study sought explanation from the testimony of principals. In Chapter 3, I present a methodology that allowed for the exploration of testimony. By examination of the recollections of principals, this study reveals the factors that influenced their decisions.



## **CHAPTER 3: METHODOLOGY**

### **Introduction**

The purpose of this study was to explore the perceptions of principals in rural Louisiana school districts concerning the cultural and contextual circumstances that influenced their rating of teachers on the COMPASS rubric. In this study, I examined the experiences of Louisiana principals, in rural school districts, during the first year of COMPASS implementation. I selected participants who were principals in districts in which no teachers were rated as Ineffective during classroom observations. In this chapter, I detail this study's epistemological basis and methodology.

### **Epistemological Foundation**

As a researcher, I sought to practice epistemological awareness as a means of aligning research methods and questions with the epistemological foundations of this study (Koro-Ljunberg, Yendol-Hoppey, Smith, & Hayes, 2009). This study's foundation of inquiry originated from an interpretivist perspective. Interpretivists seek to answer research questions from the perspective and knowledge of others (Leitch, Hill, & Harrison, 2009). Interpretivists challenge quantitative traditions and hold that all methods, even scientific, contain the potential for bias and complex interactions (Leitch et al., 2009). While there are several different approaches to inquiry within the interpretivist worldview, I based the methodology of this study on the constructivist approach. Constructivists view the creation of knowledge as a process by which an individual builds and makes sense of the information to which they are exposed (Cobb, 1994). Bodner (1986) described the construction process as "a process in which knowledge is both built and continually tested" (p. 6).

In terms of a methodological approach, researchers with an epistemological foundation in constructivism seek the knowledge constructed by individuals from their experience and environment. For example, researchers associate individual interviews as a method of inquiry aligned to constructivism (Koro-Ljunberg et al., 2009). Thus, I sought to align all aspects of the study's methodology with a constructivist approach, beginning with research questions framed in constructivist rhetoric, using purposeful, homogenous sampling, and gathering data through an individual interview protocol (Koro-Ljunberg et al., 2009). My relationship to this study was also influenced by the constructivist framework, resulting in a purposefully detached research, which placed emphasis on the testimony of the participants (Koro-Ljunberg et al., 2009). These methods are detailed further throughout this chapter.

### **Research Questions**

Three research questions guided this study in relation to the experiences of principals. How do principals in rural school districts in Louisiana describe:

1. The influence of the rural culture and context on their scoring decisions?
2. The procedures used by the Louisiana Department of Education to prepare and support evaluators during the implementation of COMPASS?
3. Their reasoning in whether or not they rated teachers as Ineffective?

The interview questions focused on the principals' personal experiences in making decisions on teacher effectiveness with the new COMPASS rubric. To fully address the rich experiences expected in responses to the research questions, this study utilized a constructivist approach. From the perspective of this research study, the interview data obtained represents a constructed narrative of the previous experience of principals (Silverman, 2000).

## Setting

For this study, I purposefully sampled within specific school districts as sites for participant selection (Creswell, 2007). The setting was established based on the positive skew in scores on the COMPASS rubric. Principals in 29 school districts did not identify a single teacher as Ineffective. Of further interest were districts that were performing below the state average in student performance and growth on comprehensive standardized exams. In the 2012-2013 school year, the statewide district average for students at grade level proficiency was 68% with an average growth of 1% when compared to the prior year (Education, 2013). Table 2 below is a table of districts in which principals did not report any ineffective teachers, despite academic achievement of students being lower than the state average.

Table 2: Percent of students performing proficiently in 4<sup>th</sup> and 8<sup>th</sup> grade, and the increase in proficiency between the 2011-2012 and 2012-2013 school years in districts in which principals did not rate any teachers as ineffective.

District	% of Students at Proficiency	$\Delta$ in Student Proficiency
Avoyelles	54%	.98%
Bienville	61%	3%
Caldwell	64%	-.53%
Catahoula	65%	.49%
Concordia	63%	.15%
Franklin	57%	1%
Iberville	64%	6%
Jackson	64%	1%
Red River	57%	-4%
Richland	55%	.40%
St. Martin	63%	3%
Tensas	44%	-2%
Union	51%	-3%

Six of these school districts are located in the Mississippi Delta region, which encompasses an impoverished, rural area on the eastern side of Louisiana and the western side of Mississippi (Nichols, 2004). This region is known for its high rate of poverty and rural isolation

(Nichols, 2004). Other districts, such as the Red River and Bienville school districts, are located in rural areas in the northwest region of the state. Only two of the districts, St. Martin and Iberville, are located in the southern portion of the state and both southern districts had positive academic progress, despite low levels of student achievement. Every district of interest fell under the classification of rural described in Chapters 1 and 2. Thus, the criterion for site selection was established as a rural district in the Mississippi Delta where principals did not rate any teachers as Ineffective on the COMPASS rubric.

## **Participants**

### **Selection**

The thirteen aforementioned districts represented rural districts where principals did not identify any teachers as ineffective on their job performance despite below average performance by students on state standardized tests. I then narrowed these 13 districts down to 3 districts: Franklin, Richland, and Tensas. These districts are within the same geographical area known as the Mississippi Delta, are geographically contiguous, and have student populations that collectively score at least 10 percent lower than the state average.

In regards to participant selection, Creswell (2007) recommends participants who are “accessible, willing to provide information, and distinctive for their accomplishments and ordinariness or who shed light on a specific phenomenon or issue being explored” (p. 119). To access to principals in these districts, I sent a letter that described the study to superintendents in each district of interest and requested to speak with principals. Once permission was granted, emails were sent to each principal in the district explaining the purpose of the study, ensuring confidentiality, the interview procedure, and requesting their participation (Creswell, 2007).

Interviews were scheduled and conducted with interested principals who had experience evaluating teachers using the COMPASS rubric during the 2012-2013 school year.

### **Data Collection**

Interviews with principals were conducted with a semi-structured set of questions, which contained prompts, where necessary, for additional details and richer explanations from participants (Cocek, 2012). When possible, interviews took place at the school site and were recorded through an audio recording device. Permission to record was acquired before each interview. Interviews were transcribed and analyzed through a coding process. Data collection ceased once a saturation point had been reached in regards to a consensus from the testimony of the principals.

As recommended by Creswell (2007), the interviews were conducted using an interview protocol to provide structure and consistency to the interview process. The interview process was conducted in a semi-structured manner, which provided the opportunity to ask participants to give greater detail. The interview protocol consisted of six questions, with additional probing for elaboration when necessary:

1. What are your general thoughts on the COMPASS rubric?
2. Please describe your training and support experiences in regards to the implementation of the COMPASS rubric.
3. What was your understanding of the scoring expectations put forth by the state and district?
4. Outside of the performance descriptors within the rubric and your time observing the teachers' classrooms, what factors, if any, influenced your rating decisions?

5. How did these external factors influence your rating decisions? Can you give an example?
6. What challenges did you face when given the possibility of rating a teacher as Ineffective?

The first three questions focused on the principals' understanding of and feelings towards the COMPASS rubric and their professional development experiences. These initial questions were important in examining the experiences of principals as evaluators. These personal thoughts and experiences revealed the extent to which principals underwent the high-quality rater training necessary to establish consistent and valid scoring decisions (Cantrell & Kane, 2012; Darling-Hammond et al., 2009; Wells et al., 1999). Questions four and five sought to understand factors that might have influenced the scoring decisions of rural principals. Question six sought to understand a principal's barriers in rating a teacher as Ineffective. Questions four, five, and six allowed for the expression of factors related to the culture and context of rural schools.

### **Data Analysis**

I engaged in the qualitative process of coding, in which several codes were identified and then used to organize and analyze themes within the testimonies of principals (Weston et al., 2001). A code is a "word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute" (Saldaña, 2012, p. 3). DiCicco-Bloom and Crabtree (2006) described the coding process as one in which "investigators review and identify text segments" (p. 318). Weston et al. (2001) also noted the involvement of researchers in the coding process: "how researchers see data and the meaning attributed to it is what makes data useful, interesting, and a contribution to knowledge" (p. 384). Basit's (2003) approach to the coding process detailed the use of a "provisional 'start list' of codes prior to fieldwork" (p. 145). I began with

an initial list of codes that came from “the conceptual framework, list of research questions, hypotheses, problem areas and/or key variables” (Basit, 2003, p. 145). To address the potential for bias, an external reviewer checked the themes I identified as part of the coding process prior to their use. I also recognized the potential for a generated list of codes to limit exploration of the testimonies, and thus allowed additional codes to emerge from the analysis of the data (Hewitt-Taylor, 2001).

The organization of data under codes reveals themes (Saldaña, 2012). As Saldaña (2012) explained, “a theme is an outcome of coding, categorization, and analytic reflection, not something that is, in itself, coded.” (p. 13). I transitioned from the identification of codes to the analysis of themes through a process known as classification. Creswell explained classification as “taking the text or qualitative information apart, and looking for categories, themes, or dimensions of information” (p. 153). Specifically, interview excerpts were associated with a code for discussion and analysis, which led to certain words and patterns being identified as themes (Creswell, 2007).

### **Data Quality**

Researchers have noted validity, data quality, and the rigor of research as emerging concepts within the field of qualitative research (Anfara, Brown, & Mangione, 2002). In this study, I sought to specifically address four areas of research quality: credibility, transferability, dependability, and confirmability (Anfara et al., 2002).

Without a credible relationship between study participants and me, questions about the authenticity and depth of participant responses might arise. Regarding methods to increase the credibility of the interactions between the researcher and study participants, Anfara, Brown, and Mangione (2002) recommended the use of “prolonged engagement in the field and member

checking” (p. 30). Given my previously noted experience in providing training and support to principals, I have established credibility with the research subjects as a professional practitioner in the field of teacher evaluation. Furthermore, I practiced member checking, in which each participant was allowed to view and correct transcripts of their interview. I also conducted a member check where the participants could provide feedback on my initial conclusions once the study reached the analysis phase.

As noted by Poggenpel and Myburgh (2003), “the researcher as instrument can be the greatest threat to trustworthiness in qualitative research” (p. 320). In years prior to this study, I conducted trainings on teacher evaluation in various Louisiana school districts. I also participated in policy discussions on teacher evaluation in committee meetings by the Louisiana Department of Education (LDOE) and the Board of Elementary and Secondary Education (BESE). Given the potential for bias, I held conversations with rural principals before developing the interview protocol, to ensure its proper development (Hill et al., 2005). Any principals involved in conversations during the development of the interview protocol were not considered for participation in this study. I also conducted interviews under a semi-structured interview protocol, which further reduced the bias of the interviewer by allowing the respondents to give extensive details and testimony, instead of answering questions narrowed down by the interviewer (Turner, 2010).

To address the transferability of the data, I adhered to a purposeful sampling strategy (Anfara et al., 2002). As previously noted, my methods of sampling specifically focused on principals in districts in which no teachers were rated as Ineffective. While I only focused on the testimony of a few principals, the purposeful sampling may have increased the transferability of the testimony to other principals in similar environments.



Researchers also view the coding process as an opportunity to address and establish the dependability of a qualitative study (Anfara et al., 2002). To address bias during data organization and analysis, testimony was organized by codes, and these codes were examined by a third party (Turner, 2010). Confirmability can be addressed by researchers through reflexive practices (Anfara et al., 2002). Finlay (2002) described reflexive practices in qualitative studies as “where researchers engage in explicit self-aware meta-analysis” (p. 210). Finlay (2002) further described the role of reflexivity as a standard practice in qualitative research:

Most qualitative researchers will attempt to be aware of their role in the (co)-construction of knowledge. They will try to make explicit how intersubjective elements impact on data collection and analysis in an effort to enhance the trustworthiness, transparency and accountability of their research (p. 212).

I practiced reflexivity by giving an account and acknowledgement of my background and experiences prior to this research study. Furthermore, I detailed the biases exposed during the methodological steps of member checking and interviewing the interviewer within this study’s data analysis.

### **Summary**

After selecting participants from school districts, I interviewed principals using a semi-structured interview protocol. From the literature, I identified codes relating to possible factors of influence. Then, I used these codes to organize the responses from the principals, which in turn revealed themes from the principals’ experiences with the COMPASS rubric in rural school systems. Throughout this process I sought to address the potential for bias and ensure the quality of data presented within this study. The responses of principals, the themes that emerged, and the relationship between the testimonies and this study’s research questions are explored in further chapters.

## **CHAPTER 4: RESULTS**

### **Participants**

Principals volunteered for this study after I contacted superintendents in the parishes of Tensas and Richland. Both of these parishes are located in an isolated, rural area of Northeast Louisiana known as the Mississippi Delta. Interviews were conducted during the fall of 2015. This chapter details the responses of each individual, which are organized by this study's three research questions.

Due to the homogenous nature in which I sampled participants, there were certain similarities between the participating principals. The three participants led rural schools in the Mississippi Delta. "Sarah" was an African-American woman who led an elementary school. An experienced educator, Sarah was within her first three years of serving as a principal. "John" was a Caucasian male who led a rural high school. John had over thirty years of experience and was nearing retirement. "Mary" was a Caucasian woman who led a rural elementary school. She was unique in that she had been at the same school for over twenty years as both a teacher and administrator. All participants were interviewed using the same interview protocol described in Chapter 3. After transcribing the interviews, the testimonies were coded with a pre-determined coding list informed by both literature and previous conversations in the field.

### **General Perspectives on the COMPASS Rubric**

Principals were first asked about their general opinion on the COMPASS rubric to determine if there were any underlying biases for or against the rubric. In particular, I was curious to learn if principals generally rejected the COMPASS system and its policies as they were associated with national and state initiated reform policy. Sarah responded favorably to the

initial question, describing her support for rubrics and accountability, “as long as the rubric spells out all of what it is intended to do.” Sarah also connected COMPASS with the larger policy work of accountability, and noted her support for both:

If you are doing the rubric and you are doing everything in a daily consistent manner, you don’t have to wait until it is time to be observed to be doing the rubric. They are trying to show you now, with accountability, this is what you should be doing in your classroom anyways. So I’m alright with any kind of rubric, because I think there needs to be accountability.

Sarah’s positive statements regarding the COMPASS system were tied to a philosophical support for higher accountability, even though accountability policies are driven from national reform efforts, as opposed to originating from local school systems. When compared to the other participants in this study, Sarah’s support for accountability policy sharply contrasted with the testimony of the other principals.

After this initial question, two additional questions were asked to further probe the principals’ thoughts. The first question asked how principals used the COMPASS rubric to effectively evaluate teachers. This question was intended to open up the opportunity for participants to discuss the relationship between COMPASS and their personal methods of teacher evaluation. Sarah discussed the relationship between the observer and teachers, musing that, “maybe we should have gone a step further and showed them things that would be guiding for them to get the 4s or to get the highest points.” Sarah also discussed how she trained teachers on the COMPASS rubric before evaluations by stating, “I told them, literally, use the COMPASS rubric, fold back the part you want to see, the highest part is the one you want to see.” Here, Sarah showed little reservation for the use of the COMPASS system, and instead focused on her own responsibility in assisting teachers to perform according to the new standards of job performance.

The second follow-up question challenged principals to discuss aspects of the COMPASS rubric that they would change. Sarah was adamant in her support of the COMPASS rubric and did not see the need to suggest any changes to the rubric. When asked to suggest changes to the rubric she replied, “No, because like I said, it gives you a lot of hints on how you can get to each level, whichever one you want to get to.” Thus, Sarah once again detailed her support for COMPASS as a measure of the job performance of her teachers.

The differences among Sarah, John, and Mary were distinct from the onset of the interview process. John had a positive initial statement when asked about the COMPASS rubric, claiming, “I think that it is actually well structured, that it has, I guess, the components that it evaluates are relevant to identifying types of good teaching.” However, compared to Sarah, John also gave much more detail about what he would change regarding how the rubric is used to evaluate teachers. Specifically, John noted the need for more subjectivity by evaluators:

For the administration, the observer to have some leeway in the final points of the evaluation process. So that even though the numbers indicate one thing, professional opinion would indicate something as well.

When asked to confirm, whether or not, they had intended to advocate for greater subjectivity from evaluators, John responded affirmatively. While John was polite in his disagreement with the COMPASS system, it was clear that he thought the system was imperfect. Furthermore, John’s requested change to the COMPASS system would give more power to the principals, which may only further the positive skew of the evaluation scores.

When asked about their general thoughts on the COMPASS rubric, Mary immediately pointed to her contention with the scoring expectations promoted by state trainings. Mary even detailed the impact this expectation had on her teachers:

I think it has its good points and bad points. The fact that we are told by the state that no one can score a 4, I was thinking about this, this morning, that really kind of hurt me,

because when they first came out with the rubric, my teachers were striving for that 4, and if they got the 4 that was great, but now they are being told that the 3 is the new 4, so they aren't pushing as hard.

Mary's criticism of the state's scoring expectations were a theme throughout her interview. Additionally, Mary noted objections to the time requirements of COMPASS when she stated, "I do not like the fact that I have to do this twice. I do not have that kind of time." Mary also criticized the vagueness of certain descriptors on the rubric by noting, "One of them is outcomes related to the ideas of the discipline, OK, everybody has different ideas of what that means. The way I think my kids deal with ideas, may not be the way in South Louisiana that teachers teach big ideas." Finally, Mary cited major philosophical differences between her view of early childhood education and the expectations of the COMPASS rubric:

My kids asking higher order questions, even though you give them the little sheet, the cheat sheet, they aren't really mature enough to ask the higher order questions, in the discussions, usually your more advanced kids will do that, but, um, how can you get a little one to do that?

Mary's outright contention with the COMPASS system revealed both an understanding and a rejection of the scoring expectations put forth by policymakers and communicated in professional development sessions. Mary's rejection of the scoring expectations may serve as an example of why principals in her area did not rate any teachers as Ineffective. Mary's example of the impact of the state's scoring expectations on the morale and motivation of her teachers may have led her to give higher scores to teachers, in comparison to principals who strictly followed the descriptors and scoring expectations of the COMPASS rubric.

Clearly, all three participants had varying thoughts on the COMPASS rubric. While Sarah was generally supportive of the COMPASS rubric, John and Mary had more specific criticisms. While Sarah may have generally supported accountability policies, John and Mary

advocated in their own ways for greater control for principals in the scoring decisions, and for a rejection of the scoring norms put forth by the state.

### **Training and Support During Implementation**

Participants were asked to detail their COMPASS training and their support experiences during its implementation. These support experiences might have come from the state, including statewide seminars and regional network support teams, or the local school district supervisors. Sarah described multiple state led experiences that prepared her for evaluating teachers using the COMPASS rubric:

We've gone through several, and done the activities, and did a sample where we actually went in and interviewed, and provided feedback by the Network Six team, they'll come and have different sessions.

Sarah further described the work of the local network supervisors and their hands-on approach to supporting principals, when she noted, "they will come in and even ask you if you are comfortable with going in and observing and evaluating." Here is evidence of the state's network supervisors providing direct support, mentorship, and consulting for principals in the field. Sarah's testimony indicated that she greatly valued this contact and guidance from the state. Sarah also explained that the training she received was oriented towards inter-rater reliability:

We have supervisors, who come in and observe, and get together, and say, are you seeing the same thing? Because, to be honest with you, a rubric should not be biased. In order for it to be a valid instrument with some reliability here, you should grade a person just like I'm grading them, if it exists to be a reliable instrument.

Sarah's description of the supervisors' support methods is similar to the descriptions of rater reliability training detailed in the recent literature on teacher evaluation (Cantrell & Kane, 2012; Ho & Kane, 2013). Clearly, from Sarah's testimony, there was a planned and coordinated effort by her superiors and mentors to address concerns related to the reliability of teacher

evaluation systems. Sarah also detailed the collaborative nature of her training experiences, specifically noting the relationship she had with other principals and district supervisors:

And again, with supervisors, we get to sit together, and then even though they have their own time to come in, I may ask them to come back and say, look I saw this, can you come in and see if you saw the same thing. Because then there is a next step, on your rubric, when you are trying to help the teachers to get, to improve on, whatever it is you are asking for them on that next step.

Here, Sarah described a close working relationship with those supervising the district's evaluation process. Sarah connected her role as a principal to the responsibility of coaching teachers to address performance deficiencies.

Though from a different school district than Sarah, John also described a similar training experience. Specifically, John described training experiences that involved support from both the state government and the local school district supervisors:

Initially, the training we received was through the state. And then, after that initial training, then of course there were people within the district structure that were able to mentor and support, so if we had questions we didn't have to necessarily go straight to the state; we could go to the district and get answers and support. Of course, if they didn't know, then we'd go to the state for more particular answers.

The similarities testimonies of Sarah and John were further evidence of the state government's knowledge, concern, and steps to address the positive skew and lack of scoring reliability found in Louisiana's previous evaluation systems. However, unlike Sarah, John did not give any indication of having a close working relationship or often seeking guidance from district supervisors or the state's network support team during his first year of using the COMPASS rubric.

The third participant, Mary, went through similar training experiences as Sarah and John, but, in sharp contrast, negatively described those trainings. Mary was from the same district as John and within the same state network region as Sarah, and thus should have experienced the

same training sessions and support as the other two participants. In describing her experiences, Mary found little value in the initial trainings:

We had the initial training...oh gosh, when was that, December of 2011-2012, I believe. It was really not good. The people actually told us we were the second group that they had trained; it was horrible. The training was not good at all. I walked out of there, like I'd been hit by a bus.

Mary's negative description of the training she experienced is distinct from the testimony of Sarah and John. With a negative perspective dominating her training experiences, it is questionable whether Mary was influenced by the trainings or implemented the practices presented. Furthermore, if Mary held beliefs that the state network team or district supervisors were unprepared to support her during the implementation of COMPASS, she may not have contacted them to request help, ask questions, or receive any additional support or mentorship.

One fact that was clear from all three testimonies was that principals were trained by the state, even the rural school districts; however, all three principals gave different accounts of the training. If each of these participants was from a different area of the state, the content and delivery of the trainings might have varied, however this was unlikely given the close geographical proximity of the participants. Instead, the participants represent how different individuals can have very distinct recollections regarding the same event. Additionally, the way in which these participants portrayed the training and the support they received may reflect how open they were to accepting the new culture surrounding teacher evaluation and scoring norms presented by state officials.

### **Understanding of Scoring Expectations**

Participants were asked to explain the scoring expectations that they understood from their training experiences on the COMPASS rubric. This question was important as one of the aims of this study was to determine the reasons principals avoided rating teachers as Ineffective.



The theme of preparing and reinforcing training expectations through training is prevalent in recent literature on effective teacher evaluation. It is possible that the study participants did not fully understand the scoring expectations of the state, and thus could not follow them.

Alternatively, even if principals understood the scoring expectations associated with COMPASS, other reasons may have influenced their tendencies to align scores with the positive skew of previous evaluation systems or to avoid giving teachers low ratings.

When first asked about the scoring expectations discussed in the COMPASS training sessions, Sarah explained a focus on inter-rater reliability:

For one school and their people to be scoring these teachers relatively easy, making high scores, and I come over here and I score these and be very critical, you know, the whole nine yards. I gather that they really were hoping that it would be an instrument where everybody would be scored, actually looking for high expectations, but using that rubric to totally score them by, free of bias.

Sarah then gave further details on the expectations communicated by the state by stating, “they (state trainers) said do not water it down, what you see, that is what you see.” Compared to the other participants, Sarah seemed to have the clearest recollections of the content delivered in the training sessions, and her general support and receptivity to the training may have contributed to her adoption of the state’s scoring expectations. However, Sarah also expressed doubts as to whether these expectations were appropriately implemented due to issues of inter-rater reliability. Sarah explained these doubts:

But now, I'm going to be honest with you, I don't always feel like that may be the case. Because, it's not the same thing, coming from when I was a teacher, I know there were some teachers rated Highly Effective, again. Even when I was a teacher there were two different people in the same building rating people, so here it is again, the same way you see something, may be different, so that may be just a little flaw in it.

Sarah was doubtful that other principals had the same sense of rater reliability and responsibility as she did. Perhaps through collaboration and networking, Sarah had observed different styles, methods, or perspectives of rating teachers from her fellow principals.

Clearly, both John and Mary differed in their relationship to the scoring norms, as they acknowledged the same norms as Sarah, but chose to defy these norms through different methods. During the interview, John openly described a non-verbal cultural norm around the scoring expectations behind the COMPASS rubric and noted his and his peers' disagreement with the norm. John noted, "Well, that was a point of contention I think, in our schooling. I don't know if it was ever written down, but it was highly suggested that no one gets a 4, no one scores at the top." John even explained his opposition to the scoring expectation that teachers would not be rated as Highly Effective on the rubric, "And the reason being is, in my observations, I found individuals who did score a 4 and had all the valid documentation to support it. So I didn't mind giving a 4, regardless." Through this testimony we see that John both acknowledges the state's higher scoring expectations, but chose to use the process of documenting evidence and artifacts to justify his own ratings of teachers. As described by John, this norm is interesting in light of the previously described phenomenon, where principals in rural districts seemingly avoided giving teachers Ineffective scores. While the scoring norms of the state expected teachers to not be rated as Highly Effective, John used documentation to award this category to those teachers whom he thought deserved the rating.

Furthermore, John clarified that the scoring expectations were largely the product of the local district and not the state trainers. John described a process in which the local district assumed greater responsibility over the course of COMPASS implementation:

I think the expectation was verbalized more by the district than by the state, and I think I found that to be true progressively through the process. Illustrations were given of that

this school had a very low percentage of 4 ratings, this school had a half rating of 4s, and they would compare that to this school is performing at this level, this school is performing at that level. So they were trying to make an argument of, that there was a disconnect.

One interesting phenomena here was that, during the communication of scoring expectations between state policymakers, the state's network support teams, and district supervisors, these higher scoring expectations were maintained with fidelity. However, when the scoring expectations were introduced into the local school culture and rural context, these expectations faced resistance from principals.

John also briefly described the use of data by the local district as a means of connecting COMPASS ratings with student performance. Mary also noted an explicit process of drawing connections between student scores and teacher evaluations, and discussed this process in detail; however, Mary's testimony on the subject further demonstrated her criticism of the state's expectations:

What the state told us is when you get your scores for COMPASS, everything should balance out; they should be equal to each other. If all your teachers are scoring 3s, all your kids should be scoring mastery or advanced. All your kids. Well, that doesn't make a bit of sense to me.

Both John and Mary described this action of trying to link teacher evaluations with the academic performance of students, although, while John attributed this to the local district, Mary attributed this expectation to the state. Regardless, John and Mary both experienced this expectation and disagreed with the premise that teacher evaluation should have any correlation with student performance. Mary opined that the effects of trying to tie teacher evaluation scores from a performance rubric to student performance, and the expectation of scoring rigor set forth by the state, resulted in demoralized educators:

To pull kids up, you have to teach in the 4 range. Okay, so they told us, no one can have 4. I was like, why did you put them on there if you can't have 4? I've had teachers get a 4

every time, because they had met their level of ability and drive. But they told them they can't have a 4. This is what I hear, well, if a 2 is what I get a 2 is that I will take.

The previous statement showed Mary's concern about scoring teachers in the COMPASS rubric. Consequently, if Mary viewed scoring teachers below Highly Effective as demoralizing, then she may have given teachers higher ratings to maintain morale within her workforce. Mary's testimony presents the possibility that principals defy the scoring norms presented in their training to address issues in their local school environment or organizational culture.

Thus, from the testimonies of the participants in this study, two major themes emerged relative to the scoring expectations of the COMPASS rubric. First, each of the participants understood the scoring expectations of the state policymakers that were set forth during the initial training sessions. Second, the understanding of expectations did not necessarily result in support for the expectations. The two principals with more than five years of experience gave examples of how they disagreed with the scoring expectations and alluded to their experiences in defying these expectations.

### **External Factors and Their Influence**

Principals were also asked to identify any external factors that might have influenced their rating decisions and to explain the influence of these factors as they may have avoided rating teachers as ineffective due to other factors. As previously discussed, two potential areas of factors were the organizational culture, particularly the school culture related to scoring teachers, and the local environmental context.

Sarah denied the possibility of external factors in her rating of teachers. She stated, "No, I strictly used the COMPASS, because that's bias if you allow any extras to come into play, outside of the rubric." Sarah further explained herself, noting, "If you let personal things, including what I think should be there, that's not there, then you're not allowing the instrument to

be what it is.” When pressed further for information, including probing to determine if she had heard other principals discuss external factors, Sarah repeated her denial of external factors and their influence:

No, I haven't. They've never talked about external factors. They usually just come in, and what they've done, when they come in, they've taken the actual rubric, per se, and they've taken exactly what's on the COMPASS rubric.

With Sarah's testimony on external factors, and her previous testimony of support for the COMPASS rubric, there is the possibility that Sarah was able to either use the COMPASS rubric as state policymakers intended or that any inflation of teacher ratings was a purely subconscious act. Unless she was issuing a denial of her actions, Sarah did not try to intentionally defy the scoring expectations to address other concerns that she may have had as a school leader.

Although Sarah was adamant in her statements regarding the lack of external influence during her experiences with the COMPASS rubric, John's testimony demonstrated that the evaluation process was not immune to external sources of influence. When first asked about the ratings they gave teachers, John responded by stating, “I found valid documentation to support the rating that I gave.” During the interview, I pressed John to give further details as to the meaning and nature of his documentation process. Again, John described how he used the process of note taking and documentation of teacher performance as a subtle way to justify how his scoring decisions may have differed from the scoring expectations presented at the COMPASS trainings. When pressed for further details regarding their use of documentation, John described the documentation process that they used to rate teachers:

Well, obviously, the notes that you take and the observations of what kind of interactions are happening inside the classroom, how the teacher performs according to the students, how the students interact, how the students take ownership. That tended to be a big point of observation with teachers, and teachers really worked to make sure those elements were part of their instruction.

Here, John revealed the role of his documentation process that accompanied his use of the COMPASS rubric. The documentation process itself may have been used to justify John's ratings of teachers in defiance or disagreement with descriptions of performance or scoring expectations associated with the COMPASS rubric. Thus, John may have used the documentation process to subvert the scoring expectations of state policymakers.

While John sought to provide documentation as to why his scoring decisions may have deviated from the COMPASS rubric, Mary had a different perspective on her role in the scoring process. In her testimony, Mary openly admitted to making scoring decisions that were contrary to the descriptors in the rubric, because of other important concerns she had relative to her teachers. Throughout the conversation, Mary gave specific examples of situations in which she had or would alter teacher ratings in a manner that was not aligned to the COMPASS rubric. For example, Mary described how she rated teachers higher than their performance levels on the rubric:

If I know you are a dedicated teacher and you are doing your dead level best each day, I'm going to jump you; I'm not going to lie, I'm not going to do that, but if I see you or one kid do it, I'm going to give it to you. Well, and they (the state) might want to see several kids doing it before you give them that Highly Effective.

Mary openly admitted to easing the requirements on the COMPASS rubric when evaluating teachers. Mary's testimony revealed the inclusion of her personal and professional assessment of a teacher, beyond that of the evaluation rubric. This clearly illustrates a principal allowing external factors to influence their scoring decisions. Furthermore, this testimony demonstrates the potential role of the local school culture in determining a principal's scoring decisions. Specifically, the culture within the school, including the relationships between Mary and the teachers she managed, influenced Mary to make decisions contrary to the rigorous scoring expectations of state policymakers.

However, according to her testimony, Mary's scoring decisions were not just positively influenced by her personal relationships with her school's staff. Mary also described situations where she expected more of some teachers, and may have rated them at a higher standard than that described by the COMPASS rubric:

I've got a couple that just don't give a flying foot if the sun comes up. No, I'm going to expect a little bit more out of you and you can give me a little bit more. Right or wrong, that's just the way that I feel.

Thus, according to Mary's testimony, while we know that she allowed the social relationships within her school's culture to influence her scoring decisions on the COMPASS rubric, it is uncertain whether or not this would have resulted in a positive or negative skew in teacher scores. However, based on the data received from the state, not a single teacher was rated as Ineffective in Mary's school district. More than likely, Mary gave higher ratings to teachers because of these social relationships and the pressures of maintaining them within her organization culture than if she strictly followed the COMPASS rubric.

Another policy component of the state's teacher evaluation system influenced Mary to positively skew teacher ratings. Mary detailed the uncertainty related to student scores and the evaluation of teachers' Student Learning Targets (SLTs), as a factor related to altering COMPASS scores:

I'm going to give your rating as I can, because we just don't know what those kids are going to earn, especially until about mid-year when we do that mid-year evaluation; we really just don't know what those kids are going to do.

Again, Mary gave a reason as to why she inflated teacher ratings on the COMPASS rubric. Here, Mary attempted to protect teachers from the uncertainty of their ratings under their SLTs. Mary directly stated how the uncertainty she felt regarding predicting student performance led her to assign ratings in an effort to provide stability for her employees.

While Sarah denied the existence of external factors and John was subtler in his acknowledgments and methods in defying scoring expectations, Mary openly acknowledged factors that influenced her scoring decisions. The organizational culture of her school, including the social relationships between the school faculty and Mary as the school leader, was an important factor for Mary during the implementation of COMPASS. Regarding her professional background, Mary was fully immersed in the social culture of her school, having served as both a teacher and administrator at the same school for over twenty years. The differences in the testimony between the participants may be explained by Mary's career longevity in her school, whereas the other two participants had worked in various places within their school district.

### **Scoring Teachers as Ineffective**

Given the scoring pattern present in rural districts, cultural, contextual, or policy factors may have inhibited principals from rating teachers as Ineffective. Testimony from the participants of this study varied on the subject of positively skewing teacher performance scores, with one participant denying the practice, another subtly referencing the practice, and one participant directly confirming the practice.

As a final question and exercise, principals were asked to explore and explain why they may avoid rating teachers as Ineffective. Throughout the interview, Sarah denied any external influences guiding her rating decisions. While she admitted that she had not rated any teachers as Ineffective, Sarah explained her actions without a justification by any external factors, "I haven't rated a teacher as Ineffective yet. I didn't see any (ineffective) teachers according to the rubric." When pressed further, Sarah simply denied having any reservations to rating teachers as Ineffective and described a conflict she had with the SLT portion of the COMPASS system and the role of observations in teacher evaluation:



Now, let me restate that, there were a couple of ineffective teachers, but not based on my evaluation, my observation when I went in the room. It was because, well, when you talk about COMPASS, the overall rating comes from the evaluations plus student performance, and that is what got them Ineffective, but that is just how it is.

Despite her consistent adamant support for the COMPASS rubric, Sarah did point out the SLT portion of the COMPASS system often found teachers ineffective due to insufficient academic growth by students. Since she noted this concern, Sarah may have subconsciously inflated teacher ratings to compensate for a sense of uncertainty regarding the performance of her teachers on their SLTs. This possible subconscious action by Sarah was further supported by Mary's open admission to using her role in the rating process to compensate for the uncertainty of her teachers SLTs.

While Sarah did not openly entertain the possibility of external factors preventing her from rating teachers as Ineffective, John, however, openly discussed a major issue associated with giving Ineffective ratings to teachers:

No, there is nothing in policy, I'm sure it's just knowing the history of trying to find valid employees to fill positions, and in a small rural high school especially, the certification issue is just astronomical.

John's testimony provides two pieces of information. First, as a principal of a rural high school in the Mississippi Delta, John may have faced historic shortages of teachers, regardless of their qualifications; therefore, John may have avoided giving teachers an Ineffective rating as a way of maintaining his present workforce and to avoid issues related to finding a new teacher or any temporary substitutes. Second, John's reluctance to rate teachers as Ineffective did not come from a particular policy or component of COMPASS, but from immediate issues within his local environment. Thus, John's testimony shows the existence of factors from the environmental context and their influence on his scoring decisions as a school leader. As noted in the literature review, the recruitment and retention of teachers in rural areas is a major issue facing rural

schools (Maranto & Shuls, 2012; Nichols, 2004). When asked for more detail, John spoke of the use of remediation instead of dismissal for struggling teachers and, detailed on how he assisted teachers with improving their performance on the COMPASS rubric:

Well, obviously if the rating comes down, and I've had that experience, um, to know that, that teacher is probably going to be worthy of remediation, and in a rural area, you almost have to know that if I lose this teacher, I may not get one as good as this teacher is, so you know the task of remediation is looming over you. So in the evaluation you try to make it obvious in documentation, these are the points that need to be worked on, so you can develop a plan that can be worked on for the remediation. And hopefully, salvage a teacher and not have to go through another employment process, where you may have a worse situation than you already have.

A picture of John's interactions within the COMPASS framework emerged in his testimony on his remediation of teachers and his previous testimony regarding documentation. It appeared that, realizing the difficulty in replacing teachers, John took specific notes and documentation on the deficiencies of his teachers and used this documentation to support and coach his teachers in preparation for further evaluations. By putting additional work on himself as a school leader, John maintained his staff, while avoiding an open defiance of the COMPASS program. Thus, instead of an outright defiance of the policy mandate, John coached his teachers to conform their performance to the descriptors and scoring categories found on the COMPASS rubric.

Mary gave a very similar testimony in regards to the need to keep their current teachers. Both John and Mary were principals in the same rural district, however Mary's school was located closer to the region's major population center, the city of Monroe in Ouachita parish. Mary cited the proximity to a non-rural district as a concern:

I compete with Ouachita parish; we are 40 minutes from Ouachita parish, and you can go to Ouachita parish and depending on your year, make 10 grand more than you can here. They have curriculum coordinators that plan their curriculum so teachers do not have to do any planning; they have reading interventionists and different programs to help their

struggling readers, while my teachers have to give up their recess time to catch those kids up.

When pressed further on how her proximity to Ouachita parish impacted her scoring of teachers, Mary responded, “That’s me rating my teachers and keeping them here.” The preceding statement clearly indicated that Mary thought she had to deviate from the performance descriptors of the COMPASS rubric and the scoring expectations of state policymakers. Unlike John, Mary did not try to work within the system; she openly made the choice to defy policy to address an immediate concern in her work environment.

### **Summary**

Within these three testimonies, Sarah stood out as the most supportive of the COMPASS rubric, even though she was from a more rural, isolated district compared to John and Mary. Sarah may have been more influenced by the culture of accountability and high expectations set forth by the state than the more veteran principals John and Mary. Alternatively, Sarah did not rate any teacher as Ineffective, and may have been unintentionally or subconsciously skewing the scores she gave teachers. Both John and Mary openly discussed their actions and conflicts during the implementation of the COMPASS rubric.

From John’s and Mary’s testimonies arises the possibility that principals may engage in subversive or defiant behavior during the evaluation process to address certain factors within their environment, particularly the retention of teaching staff in a rural school district. During the implementation of COMPASS, an Ineffective rating meant that a teacher would lose tenure and be subjected to further performance reviews. Furthermore, any teacher with consecutive Ineffective ratings would go through a dismissal process. These policies may have given further impetus for rural principals to positively inflate the performance scores of their teachers, as an Effective score on COMPASS would mean continued work and benefits.

## **CHAPTER 5: DISCUSSION**

### **Summary**

The purpose of this study was to explore how cultural and contextual circumstances influenced principals in rural Louisiana school districts during their ratings of teachers on the COMPASS rubric. COMPASS was developed as a response to previous evaluation systems, which were plagued by a positive skew in their results; the vast majority of teachers scored at the highest performance levels (Darling-Hammond et al., 1983). Within this study's literature review, several potential external factors were identified that may have influenced principals' rating decisions. First, the organizational culture, including the personal relationships and espoused beliefs with a school, may have influenced principals to positively inflate teacher ratings. Second, a host of environmental factors may have also been involved; these factors included the professional development and training that principals received, and environmental factors unique to rural districts such as declining populations, low sources of school funding, and the difficulty in recruiting and retaining qualified teachers. These factors were viewed through a conceptual framework of culture and context, which allowed the acknowledgement of both social and environmental factors unique to rural schools. Furthermore, the conceptual framework guided the analysis in this chapter.

Three areas of inquiry guided this study in relation to Louisiana principals in rural school districts: the influence of the rural culture and context on scoring decisions, the procedures used to prepare and support evaluators during the implementation of COMPASS, and scoring of teachers as Ineffective. These areas of inquiry and the testimonies of rural principals are analyzed in this chapter.

## **The Influence of Rural Culture and Context**

From the testimonies of the participants in this study, both rural culture and context influenced principals, however, the participants tended to describe the influence of context rather than culture. There may be many reasons for the acknowledgement of context over culture. First, as previously defined within the framework of culture, certain aspects of culture operate on a subliminal level, which would make it difficult for participants to acknowledge the influence of culture on their decisions (Schein, 1996, 2004). Within the testimony of this study's participants, all of the principals alluded to the influence of culture. Second, contextual factors, such as the difficulty of hiring new teachers, may be more immediate and visible concerns for principals.

Culturally, Sarah's testimony focused on the culture pushed by state policymakers and her acceptance of this culture of accountability. In describing her support for accountability systems, Sarah may have been influenced by her beliefs during the teacher evaluation process. During the interview process, Sarah was adamant in her support for COMPASS and the high expectations for teachers that she developed from the COMPASS trainings. Throughout her testimony, Sarah would not address concerns or conflicts with the expectations of evaluating teachers on the COMPASS rubric. Interestingly, despite her vocal support for high expectations for teachers, Sarah claimed to have not encountered any ineffective teachers. If there was a conflict between the expectations of the state and her local culture, Sarah did not address it. Sarah's cultural conflicts may have been on a subliminal level that she could not explicitly address (Schein, 1996).

As noted in this study's literature review, the COMPASS system was a response to prior evaluation systems that were positively skewed. However, some principals rejected the scoring paradigms associated with COMPASS. Both John and Mary acknowledged the high scoring

expectations presented by the state during training sessions. While John was subtle in his rejection of the state's scoring expectations, Mary vocally and adamantly rejected the expectation of tying teacher scores to student performance.

In Mary's case, the culture of scoring expectations pushed by state policymakers conflicted with the local culture of her school, particularly her expectations for teachers as a school leader. Mary rejected the culture of scoring expectations put forth by the state and spoke positively about her school's culture. Because of her involvement with the same school for two decades, her relationships took precedence over aligning her scoring decisions with the state's expectations.

John further elaborated on the conflicts facing him as a school leader. For example, he openly acknowledged and rejected the new culture of scoring expectations. When questioned further, John expressed concerns regarding the ability to hire new teachers if they followed the high scoring expectations of the district. When culture and context conflicted, John chose to place priority on local context rather than culture. Thus, influences of the local culture, the state culture regarding scoring norms, and local rural context presented conflicting influences on principals.

The influence of culture and context on the rating decisions of principals is evident in the testimonies of principals; however, there is not a simple or direct relationship between these factors and the decisions of principals. Although all participants cited the cultural influence of high scoring expectations pushed in state trainings, only two of those principals openly rejected this cultural imposition. In relation to this study's conceptual framework and working definition taken from Schein's (2004) work, the local culture and the state culture operate at two different levels. The local culture surrounding scoring expectations and social support for retaining local

teachers seemed to be internalized as an underlying assumption by John and Mary (Schein, 2004). However, John and Mary did not internalize the scoring expectations presented at the state-led training sessions and instead viewed them on the artifact level, where they can be openly questioned and critiqued (Schein, 2004). Sarah, however, seemed to have accepted the state's expectations, given her unquestioning and vocal support for accountability policies. For Sarah, the expectations espoused from the culture of higher accountability seemed to have been internalized and accepted as more than external artifacts (Schein, 2004). This acceptance of the state's scoring expectations may be due to Sarah's first experiences with the COMPASS rubric as a principal. Without years of cultural expectations surrounding prior teacher evaluation systems, and a relatively short tenure as a principal compared to John and Mary, Sarah may have been able to more readily accept the new scoring culture associated with the COMPASS system.

Contextually, both John and Mary cited the difficulty recruiting and retaining qualified teachers as an environmental factor. Mary cited the specific contextual concern of the loss of teachers to the nearest non-rural parish of Ouachita. John spoke more generally on the issue, citing concerns regarding his overall difficulty in finding certified teachers. These workforce issues can be the result of other environmental factors present in rural districts, such as the geographic isolation, declining population, and declining school funding. The juxtaposition between the veteran principals and Sarah was striking regarding the testimony related to contextual factors. Although Sarah was in a school district that was more rural, more isolated, and more impoverished, she did not cite any environmental concerns. In her testimony, the local context was not cited as a consideration, and she focused on her vocal support for teacher accountability. Furthermore, Sarah explicitly denied the consideration of any external factors while she was rating teachers.

Thus, for John and Mary, local culture and context took precedent over the new culture surrounding teacher evaluation. Sarah, however, seemed to directly accept the scoring expectations and culture of accountability associated with the COMPASS system, and did not provide testimony on the influence of contextual factors. The difference in the number of years of leadership among the participants may explain how Sarah was able to so readily accept the scoring expectations and cultural norms of the COMPASS system.

### **Support and Training for Evaluators**

From previous research on teacher evaluation systems, proper support and training for evaluators was identified as an important step in ensuring the validity and reliability of the evaluation process (Cantrell & Kane, 2012; Ho & Kane, 2013). Collaborative trainings, in which principals practiced evaluating sample classroom videos and discussed the scoring results of those observations, have been shown to increase the reliability and validity of evaluation instruments (Cantrell & Kane, 2012). This format of collaborative training was also described in Sarah's testimony. Sarah described how she worked with supervisors in her district and the state network support team that administered the COMPASS trainings. John and Mary also noted that they took part in training exercises led by employees from both their local school district and the state government. Since the three participants were from the same region, they would have attended the same trainings sponsored by the state.

Thus, the scoring distributions from the first year of COMPASS training were not a result of an absence of training. All participants noted various levels of training and support from sources at their local district office, regional network teams, and from the Louisiana Department of Education (LDOE). Furthermore, the messages from all levels of training promoted the same



cultural norm of high expectations when assigning scores to teachers. All participants in this study specifically noted the communication of high expectations from the training sessions.

In their testimony, principals also described the quality of the training sessions. As detailed in this study's literature review, effective professional development experiences share common traits such as being ongoing and collaborative (Garet et al., 2001). The training sessions described were clearly collaborative in their approach, involved teams of observers, and had an extended duration. The professional development provided during COMPASS implementation also took the format of a common and effective method of principal training: mentorship (Alsbury & Hackmann, 2006). For example, Sarah specifically detailed a relationship of mentorship between principals and their supervisors during the observation process. While John briefly mentioned the availability of mentorship from experienced evaluators during the implementation of COMPASS, there was no indication from their testimony that either he or Mary took advantage of this resource.

Even though the participants received training to prepare them for their role as evaluators, they may have been unable to validly and reliably rate teachers because of the quality of the professional development that they received. Although all principals were trained on the rubric and how to rate teachers at various levels based on their performance, Sarah and John provided most of the testimony regarding the quality of the training. Their descriptions of their training experiences directly aligned with noted aspects of effective professional development. For example, Sarah's testimony described multiple, ongoing training experiences in the form of both mentorships and training sessions (Grodzki, 2011; Ho & Kane, 2013; Normore, 2007).

In terms of policy implementation, principals should have the technical capacity to rate teachers as Ineffective. However, based on John's and Mary's testimony, the positively skewed

scoring results were not an issue of lacking appropriate support, but an outright willingness to defy the scoring norms taught at these trainings. John documented teachers' performances and coached his staff through their evaluations, while Mary outright rated teachers as she felt necessary, beyond the descriptors provided by the COMPASS rubric. In Sarah's case, instead of an outright defiance, there may have been a more subtle denial, as she claimed to not see any teachers that she would have identified as ineffective.

Regardless of their training experiences, none of the participants in this study had ever rated any teachers as Ineffective. The trainings were intended to promote a culture of accountability and high expectations for teacher performance, with an expected end result of more sparing use of higher ratings for teachers. However, since the training experiences provided to principals are just one of many factors relevant to principals, perhaps principals prioritized other factors. These other factors could have been of such immediate concern to principals that they would have given a teacher an undeserved higher rating. For example, John and Mary repeatedly cited concerns regarding the recruitment and retention of teachers in a rural school district.

### **Rating Teachers as Ineffective**

As noted in previous chapters, the results provided by the LDOE for the first year of COMPASS implementation were positively skewed, with less than one percent of teachers statewide receiving an Ineffective rating (LDOE, 2013a). Furthermore, in many rural districts, principals did not rate a single teacher as Ineffective. The participants in this study had never rated teachers as Ineffective during their COMPASS evaluations. Each participant was asked directly if they had encountered ineffective teachers and, if so, why they had avoided giving these teachers an Ineffective rating.

When asked about the existence of ineffective teachers, Sarah initially denied seeing any teachers that she would have rated as Ineffective, but she then noted that some teachers were later found to be Ineffective based on their SLTs. This scenario is an open admission that observations by principals, even those who express vocal support for teacher accountability, may not align with measures of student learning. This situation also conflicts with the aims of state policymakers, who pushed the expectation that higher teacher ratings should correlate to academic gains by students. Sarah was further asked if there were any reasons or factors preventing her from rating teachers as Ineffective. Although she briefly cited concern for teachers rated as Ineffective under SLTs, Sarah was adamant that there was nothing preventing her from rating teachers as Ineffective during the observation process. Since culture operates within an organization at an unseen level, perhaps Sarah could not explicitly discuss all of the ways her local culture influenced her decisions.

In juxtaposition to Sara, both John and Mary detailed reasons why they did not rate teachers as Ineffective during their evaluations. John explained how he used documentation to justify his scoring decisions, especially when they may have differed from the intent or descriptors related to the COMPASS rubric. John also described the steps he took after the initial observation to ensure that low-performing teachers improved and avoided an Ineffective rating in future observations. These steps included additional coaching, mentorship, and counseling with the teachers. These additional steps can be seen as a means of protecting his teachers and their job security, by working within the COMPASS system instead of defying its implementation.

When asked about his efforts to prevent teachers from getting Ineffective ratings, John specifically cited his difficulties in finding certified teachers. As a principal of a rural high school, John struggled to find qualified and certified teachers in subject areas such as math. To

John, the extra effort associated with intervening and coaching teachers through the performance descriptions in the rubric was worthwhile compared to the time and effort trying to find and hire new teachers.

Mary openly discussed shifting ratings both up and down on the COMPASS rubric and reasons why she avoided rating teachers as Ineffective. Mary described how her previous knowledge of teachers, particularly their work ethic, influenced her rating teachers either higher or lower than how the COMPASS rubric described their performance. Furthermore, Mary discussed factors that influenced her scoring decisions. Like Sarah, Mary also cited the uncertainty of teacher performance on their SLTs as an influential factor. Mary also cited the same concerns as John, that rating teachers as Ineffective might cause them to leave her school, and she already struggled to attract and retain teachers.

Thus, none of this study's participants rated any of their teachers as Ineffective and all gave varying justifications for their actions. Although their justifications varied, several themes emerged. First, there was a subtle admission that the participants had encountered ineffective teachers, but did not rate them as such. For example, Sarah noted that some of her teachers were found to be ineffective by other means, while John described how he helped teachers who would have been ineffective, and Mary took necessary action to protect her teachers from an Ineffective rating. Second, all principals had significant motives to rate their teachers as any category higher than Ineffective. Both John and Mary discussed the difficulty of finding teachers if an Ineffective rating pressured them to leave. Ironically, while Sarah did not verbalize this same concern regarding the recruitment and retention of teachers, her school district is more rural and isolated than where John and Mary serve as principals. Whether Sarah felt free to criticize the

COMPASS system or accountability policies did not take away from the difficult rural environment in which the school she led was located.

### **Conclusion**

This study explored the testimonies of principals during their implementation of the COMPASS rubric to determine the factors that influenced their scoring decisions. The participants of this study revealed that there were cultural and contextual factors, unique to rural school districts, which influenced their rating decisions. The participants also described the training and support surrounding COMPASS implementation as a source of major cultural conflict in regards to scoring expectations. Furthermore, two of the participants described how factors influenced them to avoid scoring teachers as Ineffective and the methods by which they helped teachers avoid Ineffective ratings.

Thus, drawing from both the literature on teacher evaluation systems and the testimonies of the participants within this study, the culture surrounding the scoring expectations set forth at state trainings was not enough to overcome the local expectations for teacher performance espoused by school leaders, or the environmental context of rural schools, especially the struggle to recruit qualified teachers.

### **Suggestions for Future Research**

The findings from this study raise additional questions that may be answered by future research. First, researchers should consider the number of years of a principal's experience as a factor in future research on implementing a policy change. Each year of experience may act as additional time for the acceptance of a cultural norm, and thus experienced principals may be more resistant to change. Second, future researchers should explore other areas within Louisiana, such as urban school districts. Removing the contextual environment of a rural school and

substituting it for an urban environment may reduce the number of factors that influence principals to positively inflate COMPASS scores. In particular, if the recruitment and retention of teachers was not an arduous task, then perhaps principals would be more likely to hold their existing teachers to higher performance standards. Finally, future researchers should study how the testimonies and experiences of principals change over time. If the COMPASS system is utilized for more years, principals may adopt its scoring expectations as the new cultural norms of teacher evaluation.

### **Concluding Discussion**

The change from the use of the Louisiana Components of Effective Teaching (LCET) to the COMPASS system and its observation rubric marked a major shift in the cultural norms regarding teacher evaluation. Although training from both the state government and local districts specifically addressed the new guidelines and scoring expectations associated with COMPASS, the overall distribution of teacher evaluation scores did not change. The testimonies of principals revealed that cultural and contextual factors greatly influenced the scoring decisions made by COMPASS evaluators. Ultimately, the concerns of principals and their positions within the local school culture took precedent over the culture of higher expectations and scoring norms pushed by policymakers during the implementation of COMPASS. The testimonies concurred that there are no easy answers that would improve the implementation of teacher evaluation policies, as the concerns cited by principals are ongoing contextual and cultural issues prevalent in rural school districts. Neither policymakers nor principals should be blamed for the positive skew found in the COMPASS scores. Policymakers implemented long-term training and support for principals, but principals still proceeded with their own interpretations of the scoring expectations. Since this study examined the implementation of the COMPASS rubric, perhaps

with additional time, the scoring expectations of COMPASS will be accepted and valued by principals over other factors.

## REFERENCES

- Act No. 54, Pub. L. No. 1033 § RS 17:10.1 3881-3997 (2010).
- Alsbury, T. L., & Hackmann, D. G. (2006). Learning from Experience: initial findings of a mentoring/induction program for novice principals and superintendents. *Planning and Changing*, 37(3&4), 169-189.
- Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative analysis on stage: Making the research process more public. *Educational Researcher*, 32(7), 28-38.
- Annetta, L. A., & Dotger, S. (2006). Aligning preservice teacher basic science knowledge with INTASC I and NSTA core content standards. *Eurasia Journal of Mathematics, Science and Technology Education*, 2(2), 40-58.
- Ashton, B., & Duncan, H. E. (2012). A beginning rural principal's toolkit: A guide for success. *The Rural Educator*, 34 (1), 1-13.
- Baldwin, B. (1995). Lessons learned from Louisiana teacher evaluation experience. *Teacher Evaluation Policy: From Accountability to Professional Development*, 99-125.
- Barnett, J. H., Wills, K. C., & Kirby, P. C. (2014). Comprehensive Educator Effectiveness Models That Work: Impact of the TAP System on Student Achievement in Louisiana.
- Basit, T. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational research*, 45(2), 143-154.
- Benson, P. G., Saraph, J. V., & Schroeder, R. G. (1991). The effects of organizational context on quality management: an empirical investigation. *Management Science*, 37(9), 1107-1124.
- Berry, B. (2002). What It Means To Be a " Highly Qualified Teacher". Southeast Center for Teacher Quality. Chapel Hill, NC.
- Bloor, G., & Dawson, P. (1994). Understanding professional culture in organizational context. *Organization Studies*, 15(2), 275-295.
- Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education*, 63(10), 873.
- Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal*, 106(1), 3-20.



- Bracey, G. W. (1996). International comparisons and the condition of American education. *Educational Researcher*, 25(1), 5-11.
- Brody, J. L., Vissa, J., & Weathers, J. M. (2010). School leader professional socialization: The contribution of focused observations. *Journal of Research of Leadership Education*, 5(14), 611-651. doi: 10.1177/194277511000501401.
- Cantrell, S., & Kane, T. J. (2012). Ensuring fair and reliable measures of effective teaching:. Bill and Melinda Gates Foundation. Retrieved from [www.METproject.org](http://www.METproject.org).
- Capper, C. A. (1993). Rural community influences on effective school practices. *Journal of Educational Administration*, 31(3), 20-38.
- Cavalluzzo, L. C. (2004). Is National Board Certification an Effective Signal of Teacher Quality? *CNA Corporation*. Retrieved from [www.nbpts.org](http://www.nbpts.org).
- CCSSO. (2011). Interstate Teacher Assessment and Support Consortium (InTASC) model core teaching standards: A resource for state dialogue. Washington, DC: Council of Chief State School Officers.
- Clayton, C. (2013). Understanding Current Reforms to Evaluate Teachers: A Literature Review on Teacher Evaluation Across the Career Span.
- Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23(7), 13-20.
- Cocok, C. (2012). Exploring education through phenomenology: A review of Gloria Dall'alba's diverse approaches. *Phenomenology & Practice*, 6(1), 95-105.
- COMPASS (2013). [Request for Training Documents].
- Creswell, J. W. (2007). *Qualitative inquiry and research design* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Daley, G., & Kim, L. (2010). *A teacher evaluation system that works*. Santa Monica, CA: National Institute for Excellence in Teaching.
- Danielson, C. (1996). A framework for teaching. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2011a). The framework for effective teaching: 2011 Louisiana edition (2011 Louisiana Edition ed.). Princeton, NJ: The Danielson Group.

- Danielson, C. (2011b). The framework for teaching evaluation instrument. Princeton, NJ: The Danielson Group.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Ascd.
- Darling-Hammond, L. (1992). Reframing the school reform agenda: Developing capacity for school transformation.
- Darling-Hammond, L. (1999). *Reshaping teaching policy, preparation, and practice. Influences of the National Board for Professional Teaching Standards*. American Association of Colleges for Teacher Education, AACTE Publications, 1307 New York Avenue, NW, Suite 300, Washington, DC 20005-4701.
- Darling-Hammond, L. (2006). Assessing teacher education the usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, 57(2), 120-138.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappa International*, 93(6), 8-15.
- Darling-Hammond, L., & Berry, B. (2006). Highly qualified teachers for all. *Educational Leadership*, 64(3), 14.
- Darling-Hammond, L., & Marks, E. L. (1983). *The new federalism in education: State responses to the 1981 Education Consolidation and Improvement Act*: ERIC.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Dallas, TX: Stanford University.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Deal, T. E., & Peterson, K. D. (1999). *Shaping school culture: The heart of leadership*. San Francisco, CA: Jossey-Bass.
- Debray, E. H., McDermott, K. A., & Wohlstetter, P. (2005). Introduction to the special issue on federalism reconsidered: The case of the No Child Left Behind Act. *Peabody Journal of Education*, 80(2), 1-18.
- Denison, D. R. (1990). *Corporate culture and organizational effectiveness*. New York, NY: John Wiley & Sons, Inc.

- Denison, D. R., & Spreitzer, G. M. (1991). Organizational culture and organizational development: A competing values approach. *Research in Organizational Change and Development*, 5(1), 1-21.
- DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, 40(4), 314-321.
- Ellett, C. D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of personnel evaluation in education*, 17(1), 101-128.
- Elmore, R. F. (2008). *School reform from the inside out: Policy, practice and performance* (5th ed.). Cambridge, MA.
- Elsberry, C. C., & Bishop, H. L. (1993). *Perceptions of first-year elementary principals in three southeastern states regarding principal induction programs*. Paper presented at the Annual meeting of the Midsouth Educational Research Association, New Orleans, LA.
- Finlay, L. (2002). Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative research*, 2(2), 209-230.
- Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). Highlights from PISA 2009: Performance of US 15-year-old students in reading, mathematics, and science literacy in an international context. NCES 2011-004. *National Center for Education Statistics*.
- Gallagher, H. A. (2004). Vaughn elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79-107.
- Gardner, D. P. (1983). A nation at risk. *Washington, DC: The National Commission on Excellence in Education, US Department of Education*.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- Ghoshal, S., & Bartlett, C. A. (1994). Linking organizational context and managerial action: The dimensions of quality of management. *Strategic Management Journal*, 15(S2), 91-112.
- Glazerman, S., McKie, A., & Carey, N. (2009). An evaluation of the teacher advancement program (TAP) in Chicago: Year one impact report. Final Report. *Mathematica Policy Research, Inc*.
- Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. National Comprehensive Center for Teacher Quality. Washington DC.

- Goertz, M. E. (2005). Implementing the no child left behind act: Challenges for the states. *Peabody Journal of Education*, 80(2), 73-89.
- Goodpaster, K. P. S., Adedokun, O. A., & Weaver, G. C. (2012). Teachers' perceptions of rural STEM teaching: Implications for rural teacher retention. *The Rural Educator*, 33(3), 9-22.
- Gordon, D. (2002). Putting National Board certification to the test. *Harvard Education Letter: Research Online*.
- Grodzski, J. S. (2011). Role identity: At the intersection of organizational socialization and individual sensemaking of new principals and vice-principals. *Canadian Journal of Educational Administration and Policy*, 127.
- Guskey, T., & Sparks, D. (2002). *Linking professional development to improvements in student learning*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Guskey, T. R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6) 45-51.
- Guthrie, J. W., & Springer, M. G. (2004). A nation at risk revisited: Did "wrong" reasoning result in "right" results? At what cost? *Peabody Journal of Education*, 79(1), 7-35.
- Hallgren, K., James-Burdumy, S., & Perez-Johnson, I. (2014). State Requirements for Teacher Evaluation Policies Promoted by Race to the Top. NCEE Evaluation Brief. NCEE 2014-4016. *National Center for Education Evaluation and Regional Assistance*.
- Hamilton, J. (2010). Delaware and Tennessee Win First Race to the Top Grants [Press release].
- Hatch, M. J. (2012). *Organization theory: modern, symbolic and postmodern perspectives*. New York, NY: Oxford University Press.
- Hazi, H. M., & Rucinski, D. A. (2009). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, 17(5), 1-22.
- Heneman III, H. G., Milanowski, A., Kimball, S. M., & Odden, A. (2006). Standards-Based Teacher Evaluation as a Foundation for Knowledge-and Skill-Based Pay. CPRE Policy Briefs. RB-45. *Consortium for Policy Research in Education*. Philadelphia, PA.
- Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2013). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record* 116(1), 1-28.

- Hewitt-Taylor, J. (2001). Use of constant comparative analysis in qualitative research. *Nursing Standard, 15*(42), 39.
- Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess, S. A., & Ladany, N. (2005). Consensual qualitative research: An update. *Journal of Counseling Psychology, 52*(2), 196-205.
- Ho, S. (2003). Accomplishment and challenges of Hong Kong education system: What we have learned from PISA. *Educational Journal, 31*(2), 1-30.
- Ho, S. & Kane, T. (2013). *The reliability of classroom observation by school personnel*. Harvard Graduate School of Education, Measuring Effective Teaching Project of the Bill and Melinda Gates Foundation.
- Honig, M. (2006a). Complexity and policy implementation. *New directions in education policy implementation: Confronting complexity*, NY: State University of New York Press.
- Honig, M. (2006b). Street-level bureaucracy revisited: Frontline district central-office administrators as boundary spanners in education policy implementation. *Educational Evaluation and Policy Analysis, 28*(4), 357-383.
- Honold, P. (2000). Culture and context: An empirical study for the development of a framework for the elicitation of cultural influence in product usage. *International Journal of Human-Computer Interaction, 12*(3-4), 327-345.
- Hord, S., Rutherford, W., Huling-Austin, L., & Hall, G. (1987). *Taking charge of change*. Austin, TX, Southwest Educational Development Laboratory.
- Howley, A., Crowley, K., & Howley, C. (2002). *Networking for the nuts and bolts: The ironies of professional development for rural principals*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. ERIC # ED 463 908.
- Hubsch, A. W. (1992). Emerging Right to Education under State Constitutional Law. *Temple Law Review, 65*, 1325.
- Hunter, J., & Back, J. (2011). Facilitating sustainable professional development through lesson study. *Mathematics Teacher Education and Development, 13*(1), 94-114.
- Ingvanson, L., & Rowe, K. (2008). Conceptualising and evaluating teacher quality: Substantive and methodological issues. *Australian Journal of Education, 52*(1), 5-35.
- James-Ward, C. (2011). The development of an infrastructure for a model of coaching principals. *The International Journal of Educational Leadership Preparation, 6*(1) n1.

- Jerald, C. D., & Van Hook, K. (2011). More than Measurement: The TAP System's Lessons Learned for Designing Better Teacher Evaluation Systems. *National Institute for Excellence in Teaching*.
- Johnson, D. (1997). Putting the cart before the horse: Parent involvement in the Improving America's Schools Act. *California Law Review*, 1757-1801.
- Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, 114 (10), 1-39.
- Johnson Jr., B. L. (1999). Great expectations but politics as usual: The rise and fall of a state-level evaluation initiative. *Journal of Personnel Evaluation in Education*, 13(4), 361-381.
- Joyce, B., & Showers, B. (2003). Student achievement through staff development. *ASCD*.
- Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research Paper. MET Project. Bill & Melinda Gates Foundation.
- Kang, N.H., & Hong, M. (2008). Achieving excellence in teacher workforce and equity in learning opportunities in South Korea. *Educational Researcher*, 37(4), 200-207.
- Kellor, E. M. (2005). Catching up with the Vaughn express: Six years of standards-based teacher evaluation and performance pay. *Education Policy Analysis Archives*, 13(7), n7.
- Kilmann, R. H. (2011). *Quantum organizations: A new paradigm for achieving organizational success and personal meaning*. Newport Coast, CA: Kilmann Diagnostics.
- Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3-13.
- Kim, S., & Lee, H. (2006). The impact of organizational context and information technology on employee knowledge-sharing capabilities. *Public Administration Review*, 66(3), 370-385.
- Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78.
- Knutson, K. M., Krueger, F., Koenigs, M., Hawley, A., Escobedo, J. R., Vasudeva, V., & Grafman, J. (2010). Behavioral norms for condensed moral vignettes. *Social Cognitive and Affective Neuroscience*, 5(4), 378-384.

- Koro-Ljunberg, M., Yendol-Hoppey, D., Smith, J. J., & Hayes, S. B. (2009). (E)pistemological awareness, instantiation of methods, and uniformed methodological ambiguity in qualitative research projects. *Educational Researcher*, 38(9), 687-699.
- Kupiainen, S., Hautamäki, J., & Karjalainen, T. (2009). *The Finnish education system and PISA*: Ministry of Education Helsinki.
- Ladson-Billings, G., & Darling-Hammond, L. (2000). The Validity of National Board for Professional Teaching Standards (NBPTS)/Interstate New Teacher Assessment and Support Consortium (INTASC) Assessments for Effective Urban Teachers: Findings and Implications for Assessments.
- LDOE. (2013a). *2013 COMPASS Final Report*. Baton Rouge, LA: Louisiana Department of Education. Retrieved from <http://www.louisianabelieves.com/docs/teaching/2013-compass-final-report.pdf>.
- LDOE. (2013b). *Compass: Report on the status of Louisiana's educator support and evaluation system and the value-added assessment model*. Retrieved from [http://www.boarddocs.com/la/bese/Board.nsf/files/95DPP36519C4/\\$file/AF\\_3-1\\_VAM\\_Summary\\_Report\\_March2013.pdf](http://www.boarddocs.com/la/bese/Board.nsf/files/95DPP36519C4/$file/AF_3-1_VAM_Summary_Report_March2013.pdf).
- LDOE. (2014). COMPASS. Retrieved 8-1-2014, from <http://www.louisianabelieves.com/teaching/compass>.
- LDOE. (n.d.). Louisiana components of effective teaching. Retrieved 9/14/2014, 2014, from <http://www.louisianaschools.net/lde/uploads/5564.pdf>.
- Le Tendre, M. J. (1996). The new improving America's schools act and Title I. *Journal of Education for Students Placed At Risk*, 1(1), 5-8.
- Lee, J., & McIntire, W. G. (1999). *Understanding Rural Student Achievement: Identifying Instructional and Organizational Differences between Rural and Nonrural Schools*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Leitch, C. M., Hill, F. M., & Harrison, R. T. (2009). The philosophy and practice of interpretivist research in entrepreneurship: Quality, validation, and trust. *Organizational Research Methods*.
- Leonard, B. (1994). *Improving America's Schools Act of 1994*: Diane Publishing.
- Lindahl, R. A. (2006). The role of organizational climate and culture in the school improvement process: A review of the knowledge base. *Education Leadership Review*, 7(1), 19-29.

- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3-16.
- Liu, S., & Teddlie, C. (2003). The ongoing development of teacher evaluation and curriculum reform in the People's Republic of China. *Journal of Personnel Evaluation in Education*, 17(3), 243-261.
- Liu, S., & Teddlie, C. (2005). A follow-up study on teacher evaluation in China: Historical analysis and latest trends. *Journal of Personnel Evaluation in Education*, 18(4), 253-272.
- Louis, K. S. (2006). Changing the culture of schools: Professional community, organizational learning, and trust. *Journal of School Leadership*, 16(5), 477.
- Manna, P., & Ryan, L. L. (2011). Competitive grants and educational federalism: President Obama's race to the top program in theory and practice. *Publius: The Journal of Federalism*, 41(3), 522-546.
- Maranto, R., & Shuls, J. V. (2012). How do we get them on the farm? Efforts to improve rural teacher recruitment and retention in Arkansas. *The Rural Educator*, 41(1), 32-40.
- McDonnell, L. M. (2005). No Child Left Behind and the federal role in education: Evolution or revolution? *Peabody Journal of Education*, 80(2), 19-38.
- McGreal, T. L. (1983). *Successful Teacher Evaluation*. Publications, Association for Supervision and Curriculum Development, 225 North Washington Street, Alexandria, VA.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171-178.
- Milanowski, A. T., Kimball, S. M., & White, B. (2004). The Relationship Between Standards-Based Teacher Evaluation Scores and Student Achievement: Replication and Extensions at Three Sites Consortium for Policy Research in Education (CPRE)-University of Wisconsin Working Paper Series. *TC*, 4(01).
- Morton, C., & Harmon, H. L. (2011). Challenges and sustainability practices of frontier schools in Montana. *The Rural Educator*, 33(1).
- Mullis, I. V. (1993). *Executive Summary of the NAEP 1992 Reading Report Card for the Nation and the States. Data from the National and Trial State Assessments*: ERIC.
- Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*: ERIC.



- Mullis, I. V., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*: ERIC.
- Nichols, J. (2004). Recruiting and interviewing teachers in rural school districts: Protocol or potluck. *The Rural Educator*, 26(1), 40-47.
- Normore, A. (2007). A continuum approach for developing school leaders in an urban district. *Journal of Research of Leadership Education*, 2(3), 1-45.
- O'Toole, L. J. (2000). Research on policy implementation: Assessment and prospects. *Journal of Public Administration Research and Theory*, 10(2), 263-288.
- Odden, A. (2004). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education*, 79(4), 126-137.
- Orton, J. D., & Weick, K. E. (1990). Loosely coupled systems: A reconceptualization. *Academy of Management Review*, 15(2), 203-223.
- Pecheone, R. L., & Chung, R. R. (2006). Evidence in Teacher Education The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education*, 57(1), 22-36.
- Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, 26(4), 909.
- Poggenpoel, M., & Myburgh, C. (2003). The researcher as research instrument in educational research: A possible threat to trustworthiness? (A: Research\_instrument). *Education*, 124(2), 418-421.
- Popham, W. J. (1972). California's precedent-setting teacher evaluation law. *Educational Researcher*, 1(7), 13-15.
- Provasnik, S., Ramani, A. K., Coleman, M. M., Gilbertson, L., Herring, W., & Xie, Q. (2007). *Status of education in rural America*. Institute of Education Sciences: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubs2007/2007040.pdf>.
- Reynolds, D., Muijs, D., & Treharne, D. (2003). Teacher evaluation and teacher effectiveness in the United Kingdom. *Journal of Personnel Evaluation in Education*, 17(1), 83-100.
- Saldaña, J. (2012). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage Publications.
- Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57-67.

- Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation. Research Report*: ERIC.
- Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, Quasi-experimental study of title 1 schools. *American Educational Research Journal*, 46. doi: 10.3102/0002831209333185.
- Schacter, J., & Thum, Y. M. (2005). TAPping into high quality teachers: Preliminary results from the teacher advancement program comprehensive school reform. *School Effectiveness and School Improvement*, 16(3), 327-353.
- Schein, E. H. (1996). Culture: The missing concept in organization studies. *Administrative Science Quarterly*, 229-240.
- Schein, E. H. (2004). *Organizational culture and leadership* (3rd ed.). San Francisco, CA: Jossey-Bass.
- Schleicher, A. (2009). Securing quality and equity in education: Lessons from PISA. *Prospects*, 39(3), 251-263.
- Schmoker, M. (2012). The madness of teacher evaluation frameworks. *The Phi Delta Kappan*, 93(8), 70-71.
- Seo, K. (2012). Lessons from Korea. *Educational Leadership*, 70(3), 75-78.
- Shakman, K., Riordan, J., Sanchez, M. T., Cook, K. D., Fournier, R., & Brett, J. (2012). An examination of performance-based teacher evaluation systems in five states. Issues & Answers. REL 2012-No. 129. *Issues & Answers Report*, 129, 32.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). Teacher evaluation: Guide to effective practice. (Vol. 41). Springer Science & Business Media.
- Silverman, D. (2000). *Doing qualitative research: A practical handbook*. Thousand Oaks, CA: Sage Publications.
- Simola, H. (2005). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education*, 41(4), 455-470.
- Smircich, L. (1983). Concepts of culture and organizational analysis. *Administrative Science Quarterly*, 339-358.
- Smith, T. W., Gordon, B., Colby, S. A., & Wang, J. (2005). *An examination of the relationship between depth of student learning and National Board Certification status*. Appalachian State University, Office for Research on Teaching.

- Solmon, L., White, J. T., Cohen, D., & Woo, D. (2007). The effectiveness of the teacher advancement program. *Santa Monica, CA: National Institute for Excellence in Teaching.*
- Stacey, K. (2012). *The international assessment of mathematical literacy: PISA 2012 framework and items*. Paper presented at the Proceedings of The 12th International Congress on Mathematical Education.
- Stelmach, B. L. (2011). A synthesis of international rural education issues and responses. *The Rural Educator, 32*(2), 32-42.
- Superfine, B. M., Gottlieb, J. J., & Smylie, M. A. (2012). The expanding federal role in teacher workforce policy. *Educational Policy, 26*(1), 58-78.
- Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly, 46*. doi: 10.1177/1094670509353043.
- Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review, 273-286*.
- Szymanski, J. (2010). *The State of Education in Louisiana: Trend Analyses of Student Achievement, Student Matriculation, and Student Behavior in the Accountability Era*. Louisiana State University. Retrieved from <http://etd.lsu.edu>.
- Teddlie, C., Ellett, C., & Naik, N. (1990). *A study of the generalizability of the system for teaching and learning assessment and review (STAR)*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.
- Theobald, P., & Nachtigal, P. (1995). Culture, Community, and the Promise of Rural Education. *Phi Delta Kappan, 77*, 132-135.
- Toch, T. (2008). Fixing teacher evaluation. *Educational Leadership, 66*(2), 32-37.
- Turner, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report, 15*(3), 754-760.
- USDOE. (2009). *Race to the Top: Executive Summary*. Washington, D.C.: Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- Voyles, M. M. (2012). Perceived needs of at-risk families in a small town: Implications for full-service community schools. *School Community Journal, 22*(2), 31-66.
- Wake, D. G. (2012). Exploring rural contexts with digital storytelling. *The Rural Educator, 33*(3), 23-37.
- Wargo, M. J. (1972). ESEA Title I: A Reanalysis and Synthesis of Evaluation Data from Fiscal Year 1965 Through 1970. Final Report.

- Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the United States and abroad. Dallas, TX: Stanford University.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 1-19.
- Weick, K. E. (1995). Sensemaking in organizations. Sage.
- Weiss, E. M., & Weiss, S. G. (1998). New Directions in Teacher Evaluation. ERIC Digest.
- Wells, Z., Rinehart, J. S., & Scollay, S. J. (1999). *Kentucky's induction model intern assistant principals: Players or designated hitters in high stakes accountability?* Paper presented at the Annual Meeting of the University Council for Educational Administration, Minneapolis, MN.
- Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman, C., & Beauchamp, C. (2001). Analyzing interview data: The development and evolution of a coding system. *Qualitative Sociology*, 24(3), 381-400.
- White, R. E. (1986). Generic business strategies, organizational context and performance: An empirical investigation. *Strategic Management Journal*, 7(3), 217-231.
- Whiteman, R. S., Shi, D., & Plucker, J. A. (2011). Revamping the teacher evaluation process. education policy brief. Volume 9, Number 4, Fall 2011. Center for Evaluation and Education Policy, Indiana University.
- Wickert, D. M. (1987). Using teacher evaluation for improving instruction. *The Clearing House*, 61(1), 23-24.
- Wise, A. E. (1984). Case studies for teacher evaluation: A study of effective practices. Retrieved from ERIC database. ( ED251952).
- Xue, L. (2006). Actuality and expectation of Korea teacher evaluation system [J]. *World Education Information*, 10, 024.
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teachers professional development affects student achievement *Issues and Answers Report, REL 2007-No. 033*. Washington DC: US Department of Education.

## **VITAE**

Keith Michael Courville was born in New Iberia, Louisiana. After obtaining a degree in Chemistry and conducting biomedical experiments for three years, he entered into the profession of teaching. After three years of teaching and coaching at Walker High School, Keith joined the staff of the Associated Professional Educators of Louisiana (A+PEL). He currently serves as the Executive Director of the Associated Professional Educators of Louisiana and anticipates graduating in May of 2016.