The Pennsylvania State University
College of Education
Department of Education Policy Studies

INFLUENCES ON PRINCIPAL EVALUATIONS OF TEACHER PERFORMANCE

A Dissertation in
Educational Leadership
by
Thomas J. Dodd

Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
August 2015
The dissertation of Thomas J. Dodd was reviewed and approved* by the following:

Nona Prestine
Professor of Education
Dissertation Advisor
Committee Co-Chair

Ed Fuller
Associate Professor of Education
Committee Co-Chair

Jacqueline Stefkovich
Professor of Education

Davin Carr-Chellman
Assistant Professor of Education

Gerald LeTendre
Professor of Education
Department Head, Education Policy Studies

*Signatures are on file in the Graduate School.
Abstract

Colorado Senate Bill 10-191, the Great Teachers and Leaders Bill (SB 191) passed in 2010 in support of the U.S. Department of Education’s Race to the Top program and No Child Left Behind Act, mandated increased teacher and principal accountability by affirming improved teacher evaluation will advance K-12 schooling. As a result, the 27,000-student Pioneer Valley School District (PVSD) in Colorado initiated a new evaluation protocol weighting 50% of teacher and principal summative performance ratings on professional practice using a new instrument called the 5 Dimensions+ Teacher Evaluation Rubric (5D+), and 50% on student achievement data using a PVSD-created value-added model (VAM). Because school principals evaluate teachers this paper uses an embedded case study and Ajzen’s Theory of Planned Behavior (2006) as the conceptual framework to simultaneously consider how attitude toward a behavior (behavioral beliefs), social pressure (normative beliefs), and behavioral controls (control beliefs) influence principals’ decisions when evaluating teachers within the context of conflicting social forces and the collegial interactions which may contribute to or challenge relational trust and the status quo. While the content of teacher evaluations written by building principals is typically invisible to outsiders due to the confidential nature of personnel files, the ramifications of evaluation decisions are keenly felt by school and district insiders, embedded in the broader educational environment, and inherently multilevel. The purpose of this study then is to build an analytical hypothesis using descriptive data from in-depth interviews and document analysis to examine principals’ perceptions, interpretations, and actions in the context of teacher evaluation that can be refined and tested in future empirical research.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>vii</td>
</tr>
<tr>
<td><strong>Chapter 1: Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>Problem Statement</td>
<td>1</td>
</tr>
<tr>
<td>Research Questions</td>
<td>11</td>
</tr>
<tr>
<td>Conceptual Framework</td>
<td>12</td>
</tr>
<tr>
<td>Significance of this Study</td>
<td>14</td>
</tr>
<tr>
<td><strong>Chapter 2: Literature Review</strong></td>
<td>17</td>
</tr>
<tr>
<td>History of Evaluation and Accountability</td>
<td>17</td>
</tr>
<tr>
<td>Teacher Effectiveness</td>
<td>22</td>
</tr>
<tr>
<td>The Principal’s Role in Effective Teacher Evaluation</td>
<td>27</td>
</tr>
<tr>
<td>Sensemaking</td>
<td>29</td>
</tr>
<tr>
<td>Principal Decision-making in Teacher Performance</td>
<td>31</td>
</tr>
<tr>
<td>Confronting Poor Teaching</td>
<td>35</td>
</tr>
<tr>
<td><strong>Chapter 3: Research Methodology</strong></td>
<td>39</td>
</tr>
<tr>
<td>Justification for a Qualitative Approach</td>
<td>40</td>
</tr>
<tr>
<td>Rationale for a Case Study</td>
<td>42</td>
</tr>
<tr>
<td>Site Selection and Principal Recruitment</td>
<td>42</td>
</tr>
<tr>
<td>Confidentiality and Anonymity</td>
<td>44</td>
</tr>
<tr>
<td>Role of the Researcher as a Participant Observer</td>
<td>45</td>
</tr>
<tr>
<td>Researcher Identity</td>
<td>47</td>
</tr>
<tr>
<td>Data Collection</td>
<td>49</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>51</td>
</tr>
<tr>
<td>Credibility and Trustworthiness</td>
<td>53</td>
</tr>
<tr>
<td>Limitations of a Case Study Design</td>
<td>55</td>
</tr>
<tr>
<td><strong>Chapter 4: Study Findings</strong></td>
<td>57</td>
</tr>
<tr>
<td>Perceived Factors Influencing Principals’ Evaluations of Teachers</td>
<td>58</td>
</tr>
<tr>
<td>Principal-Teacher Relationships</td>
<td>58</td>
</tr>
<tr>
<td>Individualizing evaluations</td>
<td>58</td>
</tr>
<tr>
<td>Probationary versus non-probationary</td>
<td>64</td>
</tr>
<tr>
<td>School Level Factors</td>
<td>67</td>
</tr>
<tr>
<td>Workload</td>
<td>67</td>
</tr>
<tr>
<td>Climate and culture</td>
<td>71</td>
</tr>
<tr>
<td>Parent perceptions and feedback</td>
<td>73</td>
</tr>
<tr>
<td>District Level Factors</td>
<td>74</td>
</tr>
<tr>
<td>5D+ evaluation rubric</td>
<td>75</td>
</tr>
<tr>
<td>Professional development support</td>
<td>76</td>
</tr>
<tr>
<td>Human resources (HR) and cabinet support</td>
<td>78</td>
</tr>
<tr>
<td>Instructional technology support</td>
<td>83</td>
</tr>
<tr>
<td>Research and evaluation support</td>
<td>84</td>
</tr>
</tbody>
</table>
Isolation and lack of autonomy.................................................................87

Chapter 5: Summary, Conclusions, and Recommendations...............................90
  Summary of Major Findings........................................................................92
  Conclusions............................................................................................93
    Conclusion #1.......................................................................................94
    Conclusion #2.......................................................................................96
    Conclusion #3.......................................................................................99
    Conclusion #4......................................................................................100
    Conclusion #5......................................................................................102
  Recommendations..................................................................................106

References..............................................................................................108
Appendix A: Participation Solicitation Cover Letter......................................125
Appendix B: Signed Consent Form to Participate in a Social Science Research Study...........126
Appendix C: Interview Questions..............................................................128
Appendix D: Principal and School Demographic Information........................129
Appendix E: Glossary................................................................................130
Appendix F: PVSD Teacher Evaluation Process Overview............................131
Appendix G: PVSD-designed Value-added Measure (VAM)............................132
Appendix H: PVSD Shared Educator Growth Rating Calculation....................133
Appendix I: PVSD Teacher Rating Calculation............................................134
Appendix J: 5D+ Teacher Evaluation Rubric...............................................135
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1: Pioneer Valley School District’s SB 191 Implementation Timeline</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2: Theory of Planned Behavior</td>
<td>14</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

Completion of this study would not have been possible without the support and direction of my dissertation advisor and committee co-chair Dr. Nona Prestine, committee co-chair Dr. Ed Fuller, and committee members Dr. Jacqueline Stefkovich and Dr. Davin Carr-Chellman whose knowledge, critique, and insight helped form, refine, and improve my thinking and writing. A special thank you goes to the late Dr. Bill Boyd, who served as my initial advisor and originally gave me the opportunity to come to Penn State as a graduate research assistant, certification officer, and asst. editor of the American Journal of Education.

I also extend my sincere gratitude to the school principals and district administrators of the Pioneer Valley School District who shared their stories with me and offered the support of their respective offices to this research project.

Finally, I am most grateful to my wife and soulmate, Amy, and our two wonderful daughters, Riley and Chloe, to whom this dissertation is dedicated, for their unconditional love, understanding, patience, support, and stamina during my pursuit of the PhD. They are my motivation and my life.
Chapter 1: Introduction

Problem Statement

According to the National Council on Teacher Quality, U.S. Education Secretary Arne Duncan initiated a major shift in federal education policy in 2009 when he demanded states wishing to be considered for the 4.35 billion dollars in Race to the Top funds ensure every student had a competent teacher and principal by making student achievement a significant factor in their evaluations.

The federal Race to the Top competition used the term “effectiveness” repeatedly, rather than the “highly qualified” definition from the No Child Left Behind Act. For the first time, federal policy focused on both teachers and principals, calling for “educator effectiveness,” and it was one of the four areas where states were required to make progress to receive State Fiscal Stabilization Funds under the American Recovery and Reinvestment Act (Exstrom, 2010, para. 19).

The opening page of the Colorado Department of Education’s website (2012) titled Educator Effectiveness: A Colorado Priority stated,

Classroom teaching and school leadership are the strongest school-based factors impacting student achievement. CDE is supporting the state’s 178 school districts (to) attract, prepare and support great educators. Because... every child in every classroom deserves to have excellent teachers and excellent building leaders… (n.d., para. 1).

This statement supported the no excuses approach to school improvement, which became fashionable in the United States over the past decade, but ignored the complex social forces influencing principal practice at the building level as well as the many non-school factors affecting student performance that teachers and principals have little to no control over. The
statement spoke to the notion that if we subscribe to what former President George W. Bush famously called “the soft bigotry of low expectations,” then we are letting schools, and more specifically their teachers and principals, off the hook. Bush noted,

We are transforming our schools by raising standards and focusing on results. We are insisting on accountability, empowering parents and teachers, and making sure that local people are in charge of their schools. By testing every child, we are identifying those who need help, and we are providing a record level of funding to get them that help. In northeast Georgia, Gainesville Elementary School is mostly Hispanic and 90 percent poor, and this year 90 percent of the students passed State tests in reading and math. The principal expresses the philosophy of his school this way: “We don’t focus on what we can’t do at this school. We focus on what we can do, and we do whatever it takes to get kids across the finish line.” See, this principal is challenging the soft bigotry of low expectations [italics added]. And that is the spirit of our education reform and the commitment of our country: No dejaremos a ningun nino atras. We will leave no child behind (Bush, 2004, para. 31).

On May 12, 2010, the Colorado State Legislature passed Colorado Senate Bill 10-191 – the Great Teachers and Leaders Bill (SB 191) as a statutory effort to raise expectations, and the Pioneer Valley School District (PVSD) began planning to implement the bill locally soon thereafter (see Figure 1).
This landmark Colorado law tied teacher evaluations to the progress of students on achievement tests and principal evaluations to the performance of teachers. Colorado Senator Michael Johnston, a former principal and the primary author of the bill said, “Teachers are successful only if students are, and principals are successful only if teachers are. That’s completely opposite from how it used to be. Now we know that teachers and principals have their jobs because their students are learning” (Exstrom, 2010, para. 34). Teacher evaluations have typically served dual imperatives, which at times have been be contradictory: 1) communicate credible information about quality instruction to teachers, and 2) function as a tool for principals to document and confront poor performance. Consequently, principals, who reside at the intersection of these two conflicting narratives, are tasked with deciding in which direction to proceed, but what do they really think about their level of power and authority to do either job accurately and ethically?
This legislation was broadly consistent with legislation enacted in other states supporting a national trend to overhaul teacher evaluation and how teacher tenure was earned and maintained.

According to the National Council on Teacher Quality, 36 states and Washington, DC have reformed their teacher evaluation systems since 2009. As of November 2012, 43 states required annual teacher evaluations, 32 incorporated student achievement, 26 differentiated levels of performance, 32 mandated annual classroom observations, 22 demanded multiple classroom observations, and 9 had instituted performance-based tenure decisions (McGuinn, 2012). Requiring all Colorado school districts and Boards of Cooperative Educational Services (BOCES) to implement evaluation systems that aligned with the teacher and principal quality standards and the state board rules by July 2013, SB 191 endeavored to hold teachers and principals accountable for student learning through five core components:

1) It reformed tenure by allowing teachers to earn non-probationary status after three years of sufficient student academic growth, making non-probationary status portable from district to district within the state and revocable following two consecutive years of insufficient growth.

2) It defined an effective teacher and created parameters for an annual evaluation system that requires 50% of a teacher’s evaluation to be based on student achievement data using multiple measures.

3) It required principals to be evaluated annually with 50% of their evaluation based on student achievement and their ability to develop teacher effectiveness in their buildings.

4) It eliminated forced teacher placement (slotting teachers in schools without their or the principal’s consent) and replaced it with mutual consent hiring using the Chicago
model (principals and teachers must agree to teacher placements and teachers who are not selected serve as substitutes for a year and, if not selected in the subsequent hiring cycle, are put on unpaid leave).

5) It eliminated last in first out which allows school districts to make reduction in force decisions based on teacher performance rather than seniority (Senate Bill 191, 2010 Colo. ALS 241).

Although there was a great deal of promise and potential in this legislation, policy is only part of the answer. Even the best evaluation systems prove inadequate if poorly implemented or undermined. Researchers such as Boyer (1983), Goodlad (1984), and Sizer (1984) consistently showed that successful school reform required change to occur within the professional and political confines of the local school community level. Colorado school principals are the key leverage points in putting SB 191 into practice. After all, judgments in the face of uncertainty are the historical norm in principal decision-making. The purpose of this study was to examine principals’ perceptions, interpretations, and actions in the context of teacher evaluation using descriptive data from in-depth interviews. To date, little research has explored how principals interpret, implement, and manage these types of high accountability reforms within the context of their school buildings and larger communities. Colorado is not alone in undergoing dramatic educator evaluation system changes, as states across the country have responded to federal policy influences from President Obama’s Race to the Top program and the conditions states have to meet to receive waivers from certain provisions required by the No Child Left Behind Act (NCLB). According to a November 2012 survey of thirty-two NASSP State Coordinators, active principals or assistant principals serving as their state’s federal advocacy liaison for NASSP, teacher evaluation was their most pressing issue (NASSP, 2012).
Because research (McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997) has shown teacher quality to be the most important school-based factor affecting student achievement, non-school factors notwithstanding, policy makers have responded by legislating this responsibility on the principal. In so doing they ignored the reality that principals’ decisions about teacher effectiveness can be influenced heavily by the consequences of their decisions, the opinions and social pressure of others in their schools and districts, and their perceived level of control; not to mention that the enormous, growing, social disparities in this country also play a major role in student achievement (Ladd & Fiske, 2011). For principals, calculating teacher effectiveness is a political undertaking, increasingly driven by external forces, including the research and policy communities, and reform initiatives like SB 191. Since a major change like a high-stakes evaluation system impacts the relationship between principals and teachers, effective principal leadership requires the political skills and social power to solve conflicts and build consensus through savvy decision making (Benveniste, 1989; Willower, 1991).

McClure defined decision making as “the conversion of information into action,” which implies the central role of information (1978, p. 382). Making sense of information is a component of human behavior and critical to the decision making process (Taylor, 1986). Thus, organizational decision making is, in essence, information behavior, in which a principal’s information behavior is a product of the interaction between the principal and the environment (Rosenbaum, 1993). Bryk and Schneider’s 2002 book on Chicago school reform, *Trust in Schools: A Core Resource for Improvement* acknowledged the importance of social capital, trust, and reciprocity to effective principal leadership. Siranni and Friedland defined social capital as “those stocks of social trust, norms, and networks that people can draw upon to solve common
problems” (n. d.). Fitzpatrick (1989) reviewed research from social psychology on the effects of group discussion and exposure to other group member’s opinions. She found that group discussion may polarize preexisting opinions or cause “social comparison,” with subjective judgments the most susceptible to this influence (p. 317). Because principals’ decisions about teacher performance should exhibit a high degree of inter-rater reliability, based on the most objective information available, Fitzpatrick advocated evaluation decisions ought to minimize the effects of social comparison and maximize the usage of pertinent information. Such influences have also been discussed by Meyer and Boyd (2001) and in Goldring and Hausman’s (2000) research on the school principal’s role in community partnerships and development efforts. Stone (1998) defined the larger, related concept of “civic capacity” as the “degree to which a cross-sector coalition comes together in support of a task of community wide importance” (p. 234). Obviously, having effective teachers working with students in the local public schools is a task of community wide importance.

A wide-ranging report studying teacher evaluation in twelve diverse school districts across four states with student populations from 4,000 to 400,000, known as *The Widget Effect* (2009), illustrated the significance of having a stable of competent teachers in schools, but that teacher evaluations codified the misconception that all teachers were essentially interchangeable. More specifically, it found:

- **All teachers were rated good or great.** Less than 1 percent of teachers received unsatisfactory ratings, even in schools where students failed to meet basic academic standards, year after year.
- **Excellence went unrecognized.** When excellent ratings were the norm, truly exceptional teachers were not formally identified. Nor were they be compensated, promoted or retained.

- **Professional development was inadequate.** Almost 3 in 4 teachers did not receive any specific feedback on improving their performance in their last evaluation.

- **Novice teachers were neglected.** Low expectations for beginning teachers translated into benign neglect in the classroom and a toothless tenure process.

- **Poor performance went unaddressed.** Half of the districts studied did not dismiss a single tenured teacher for poor performance in the past five years. None dismissed more than a few each year (Weisberg, Sexton, Mulhern, & Keeling, para. overview/findings, p. 2).

Citing Jacob (2011), Reeves stated in the closing keynote address at the 2012 National Association of Secondary School Principals (NASSP) Breaking Ranks K-12 Annual Convention in Tampa, FL that evidence of principals’ ambivalence to confronting inept teaching was found in the context of Chicago Public Schools where principals had the power to non-renew (dismiss) teachers without justifying or explaining their decision to teachers per a 2004 labor agreement based on deficiencies with instruction, i.e. planning, methods, subject matter knowledge; deficiencies with environment, i.e. classroom management, teacher-pupil relationships; deficiencies with professional and personal responsibilities, i.e. attendance, tardiness, professional judgment; deficiencies with communication, i.e. parent conference skills, relations with staff; or deficiencies with attitude, i.e. lack of cooperation, lack of respect for others. However, principals of 28 – 46.2% of schools did not dismiss any teachers even in the lowest-performing schools (p. 412). Some chronic absentees and poor performing elementary teachers
were fired, but there was no relationship between achievement test scores and dismissals. The generosity of Chicago principals in their evaluations of teachers was most evident when, “Only 15 out of the 11,621 teachers who were evaluated in 2007 received a rating of unsatisfactory, and only 641 out of 11,621 (roughly 5.5 percent) received a rating of satisfactory. The remaining teachers were rated excellent or superior” (p. 406). The average dismissal rate was 11 percent, with .0013 percent rated unsatisfactory, 5.5 percent satisfactory, and the other 94 percent rated excellent or superior (Jacob, 2011).

Teacher quality was also examined in the documentary film Waiting for Superman. Eric Hanushek, an expert from CALDER, the federally funded National Research and Development Center, stated if we could replace the bottom 6 to 10 percent of teachers with average teachers, not superstars, student achievement levels in the United States would move from below average in comparison with other countries to near the top (2010). While Waiting for Superman made a case for the importance of teacher quality, it provided little insight on what influences how principals assess teacher quality. Because the findings of The Widget Effect, the Chicago study, and Waiting for Superman reflected negatively on historical evaluation practices in schools, and the dictum “past behavior is the best predictor of future behavior” has been supported by empirical evidence (Ajzen, 2011, p. 1,120), I’m was curious how principal attitudes, intentions, and behaviors toward their decision making authority regarding evaluations has been influenced by educator accountability legislation, or in the case of PVSD, SB 191. Intention, after all, is an indicator of a principal’s readiness to engage in a certain behavior.

In The Autonomy Gap Adamowski, Therriault, and Cavanna (2007) found:

- The “autonomy gap” – the difference between the amount of authority school district principals thought they needed in order to be effective leaders and the amount they
actually had – was greatest with regard to personnel decisions (i.e., the ability to
discharge or transfer unsuitable teachers, hire teachers and staff, and determine the
number and type of faculty and staff positions). These barriers were real, not
imagined, and born from state policies, district procedures, and collective bargaining
agreements.

- This autonomy gap was smaller for principals in a nonunion right-to-work state, at
  least with respect to teacher hiring, transfer, and placement decisions.
- Charter school principals felt they had greater autonomy with regard to key school
  functions than did district-operated public school principals.
- Despite the constraints they faced, most public school principals felt they had the
  ability to exercise effective leadership within the terms of their jobs. In general, they
  accepted their job for what it was, and instead of trying to change the system, they
  learned to work the system. They saw their role as “middle manager” – not CEO.
- Principals with greater longevity in their school districts felt a sense of de facto
  autonomy, as long-standing relationships allowed them to bargain for greater
  authority and school resources. These principals had honed their political and
  relationship-building skills over time.
- While principals in “managed instruction” districts – those that centrally prescribed
  curricula and instructional methods for their schools – described a lack of autonomy,
  several principals in a more decentralized district did, too. Although leaders of that
  district used rhetoric about empowering principals with greater autonomy, in practice
  union contract provisions severely limited their autonomy.
Principals in a right-to-work state felt just as restricted as principals in unionized states in discharging ineffective teachers. Enrollment laws and due process requirements were equally challenging and time consuming in both types of states (pp. 5, 6).

The report went on to say, “Nobody… sat down face-to-face with a decent sized population of principals for extended interviews so as to understand more clearly what barriers they confront and where these obstacles come from” (Adamowski, et al, 2007). This increased accountability on building principals warranted a closer look at the macro and micro forces that influenced principals’ decision making behavior as they considered evidence of teacher effectiveness and ineffectiveness when evaluating teachers, and this study endeavored to do that.

**Research Questions**

Because policymakers created a new state accountability system that changed how teacher performance is measured, how tenure is awarded or removed, how frequently evaluations occur, etc. this study identified and assessed what and how factors influenced principals’ evaluations of teachers’ professional practice (one half of a teacher evaluation) rather than how students performed on national, state, and local assessments (the other half of a teacher evaluation) within the Pioneer Valley School District. Since teacher evaluation rules initiated in Washington, DC and Denver are implemented locally, it was imperative we learn more about principal decision making and what issues they faced in assessing teacher performance in practice. Using an embedded case study design, this study focused specifically on two major questions as the investigative lens to explore and address this overarching problem:
1) What factors do principals perceive as influencing and shaping how they evaluate teachers’ professional practice?

2) How do principals perceive the manner in which these factors influence and guide their performance ratings of teachers? In what ways did the principals feel these factors influenced their performance ratings of teachers?

These two research questions were meant to serve as working guidelines, and were under continuous review and reformation as a result of the inductive reasoning used to answer them.

**Conceptual Framework**

Flowing from the constructivist/interpretivist paradigm, this study used the Theory of Planned Behavior (TPB) designed by Icek Ajzen at the University of Massachusetts to analyze principal decision making in evaluating teacher performance. According to TPB (see Figure 2) human social behavior is influenced predominantly by three variables:

- **Behavioral beliefs**- the perceived consequences of the behavior in question which generates a favorable or unfavorable attitude toward the behavior,

- **Normative beliefs**- the normative expectations and opinions of others which result in social pressure or a perceived subjective norm, and

- **Control beliefs**- the existing factors that might suppress or promote the performance of the behavior resulting in perceived behavioral controls (Ajzen, 2012a).

Collectively, these three influences, attitude toward the behavior, social pressure, and behavioral controls create a behavioral intention. Typically, a person’s intention to perform a behavior in question is strongest when the attitude toward the behavior and peer pressure is most encouraging, and the perceived behavioral control is high. Consequently, when a person has a
certain degree of control over the anticipated behavior the expectation is that person actually carries out the behavior when the opportunity presents itself. Intention therefore is a precursor of behavior completion (Ajzen, 2012a).

The TPB, a framework often used in health-related research to assess behavior (i.e. smoking cessation, and the likelihood one would engage in exercise), evolved from Bandura’s social cognitive theory. Bandura argued the importance of self-efficacy, or the belief in one’s ability to demonstrate control through motivation, affect, and action over events that affect one’s life (1989). Therefore, low self-efficacy could result in a decrease in performance. Bandura described employees with low self-efficacy this way:

People with a low sense of efficacy avoid difficult tasks. They have low aspirations and weak commitment to their goals. They turn inward on their self-doubts instead of thinking about how to perform successfully. When faced with difficult tasks, they dwell on obstacles, the consequences of failure, and their personal deficiencies. Failure makes them lose faith in themselves because they blame their own inadequacies. They slacken or give up in the face of difficulty, recover slowly from setbacks, and easily fall victim to stress and depression (1997a, p. 5).

Furthermore, Bandura (1997a) noted higher goals often motivated people to work harder to attain them, if they stayed committed over time. When challenged with unattainable goals, people typically attempted to reach them provided there was no cost of failure. This cost discrepancy was applicable to understanding how principals responded to the mandates of SB 191, which involve serious consequences for teachers and principals rated ineffective (SB 191, 2010). In addition, Bandura (1997b) found that people were likely to abandon unattainable goals when 1) they required significant effort and resources, 2) failing produced negative consequences, or 3)
other activities were available that produced positive outcomes with less effort (p. 134). What was interesting is these three factors that could result in goal abandonment are frequently present when principals are evaluating teachers.

Recognizing that people learn by observing others, the TPB framework, which added the importance of one’s social network to its roots in Bandura’s social cognitive theory, was used as the guide in exploring how different factors influenced principal decisions in evaluating teachers per SB 191. The assumption was principals’ decision making experiences, and the outcomes they promoted, were historically embedded in the broader educational environment. Therefore, individual realities were inherently multilevel, and varied across school settings (Talbert & McLaughlin, 1999).

![Figure 2: Theory of Planned Behavior diagram (Ajzen, 2006).](image)

Using the TPB as the conceptual framework, information was collected about the behavioral, normative, and control-related determinants of principals' decision-making (Ajzen, 2012a).

Significance of this Study
Even with the widely understood importance of effective principal leadership in increasing teacher efficacy through performance-based evaluation systems, few studies have looked at principals’ opinions of their level of power or authority to get the job done. Power, as a criteria in decision making (Pfeffer, 1992), impacts the organizational contextual influences on information use for decision making (O’Reilly, 1983), and relates to Ajzen’s notion of behavioral control. Principals’ willingness to provide teachers with the information necessary to enable teachers to improve their practice may be tempered by principals’ own competitive notions of power (Kirby & Bogotch, 1993). Goldstein, Marcus, and Rausch (1978) described how groups often want research to satisfy external demands, but at the same time want results to justify established policies and procedures, so decision makers are more receptive to research conclusions that fit nicely into their established policies, which speaks to Ajzen’s idea of social pressure. Taylor suggested that what administrators emphasize and reward says a great deal about the importance of different kinds of information (Pelz & Andrew, as cited in Taylor 1991). Information is more likely to be used by decision makers when it includes an effective set of incentives (O’Reilly, 1983), or what Ajzen called perceived consequences that generate a favorable attitude toward doing the behavior.

There was a pressing imperative to explore what influenced how school-based leaders inferred, analyzed, interpreted, and acted on information when evaluating teachers in this era of heightened accountability linking teacher performance to student performance. This study began this process by examining principals’ perceptions, interpretations, and actions in the context of new teacher evaluation requirements. There have been relatively few empirical studies at the local, state, or national level that explored the influences on principals’ evaluations of teachers’ professional practice. Using the TPB conceptual framework to evaluate the attitudes, social
pressure, level of behavioral control, intentions, and ultimately decision-making of PVSD principals in assessing teacher performance, the researcher followed a bottom up line of inquiry beginning with principal practice at the building-level, and working upward toward district processes and ultimately state and federal policy mandates to understand the way in which principals made sense of and reacted to local and remote influences. This study was significant because it emphasized the role of principals in putting performance accountability laws like SB 191 into practice by focusing on proximal school and school district social contexts. Better knowledge of the forces that impacted how principals perceived and intended to use information is useful to district leadership and policy makers generating hypotheses about how to best implement teacher evaluation systems as a strategy to improve K-12 education. Without this knowledge, it is difficult to conclude if the current trend of teacher effectiveness reform centered on increased teacher accountability is improving the content, process, or product of principals’ day-to-day evaluations of teachers in schools. Until principal attitudes, social pressures, behavioral controls, and intentions are recognized and examined, determining the impact of SB 191 and other such state legislative policies concerning teacher evaluation accountability cannot effectively occur (Talbert & McLaughlin, 1999).
Chapter 2: Literature Review

This review of relevant literature on teacher evaluation, decision-making, effectiveness, and accountability flowed from the problem statement and research questions to identify gaps and contradictions in the existing scholarship that merited further investigation of the behavior of principals when assessing teacher performance which held across school, district, and state contexts. For example, few published works have looked at the demand side of the teacher labor market in assessing principals’ exercise of their teacher evaluation authority (contract renewals and nonrenewals), their reasoning for making such decisions, and how principals put newly minted teacher and principal accountability policies into practice at the local level. Through this literature review the researcher added knowledge to the field and made an original intellectual contribution to answer the first research question: What factors do principals perceive as influencing and shaping how they evaluate teachers’ professional practice? Accordingly, this literature review looks at 1) the history of evaluation and accountability, 2) teacher effectiveness, 3) value-added measures, 4) principals role in teacher evaluation, 5) principals decision-making, and 6) confronting poor performance.

History of Evaluation and Accountability

In 1962 historian Raymond Callahan criticized American educational practices in his classic book entitled *Education and the Cult of Efficiency*. Reflecting on the concepts behind Frederick Taylor’s 1911 publication, *The Principles of Scientific Management*, Callahan accused American educators of adopting the philosophies and techniques of the business-industrial world and disregarding the needs of American school children. Callahan argued that,
In an effort to respond to the critiques of corporate executives and to cope with the increasingly large numbers of children pouring into schools, school leaders became obsessed with efficiency and with related ‘scientific’ practices, such as child accounting, teacher supervision, and student testing (Begley & Stefkovich, 2004, p. 132).

Although over 50 years have passed, Callahan’s criticisms are still meaningful today. Boyd (2004) referenced Callahan when noting proponents of the accountability movement, often referred to as corporate education reform, endorsed the no excuses rhetoric and furthered the standardized testing agenda to sort and select students and link teacher and principal evaluations to student scores. Today, pressures for greater efficiency, effectiveness, and accountability in public education have again become the norm in contemporary America. Yet the National Center for Education and the Economy (Tucker, 2011) didn’t find evidence that the carrot-and-stick strategy improved student achievement in the United States or any of the world’s highest performing school systems. According to Boyd, “We face a twin crisis of performance and legitimacy, as our increasingly diverse and demanding public wants more than a ‘one size fits all’ education. All of this is underscored by pressures for higher student achievement, coming from the demands of the new economy for a workforce that can ‘work smarter’ to remain competitive in the information age; and from growing pressures to close the yawning racial gap in educational achievement” (2004, p. 163).

This crisis of performance versus legitimacy is real, and often leads to discussions about educational productivity, or “the idea that more education would lead to many other social goods, such as increased economic success, greater social harmony, less poverty, less crime, and the like” (Levin, 1993, p. 1). Levin (1993) defined educational productivity as, “the search for patterns of school organization that produce the best student outcomes… in economic
terminology, the effort is to find a production function – a mathematical expression of the relationship between inputs and outputs in education” (p. 2). David Monk, a leading researcher on education production functions, noted that educational outcomes are multiple, jointly produced, and hard to compare. To measure educational productivity at the school, district, and state levels policy makers adopted state-level curriculum standards and standards-based assessments to test whether or not children met the standards (Levin, 1993). The cult of efficiency and emphasis on educational productivity in America was also evident in the form of school and district level accountability tied to the 2001 federal NCLB law, which reauthorized the former Elementary and Secondary Education Act of 1965.

Holding schools accountable through large scale state-level testing has been the pre-eminent educational reform since the mid-1990s. Stecher and Baron stated (1999), “Thirty-nine states now administer some form of performance-based assessment, 24 states attach stakes to their tests, and 40 states use test scores for school accountability purposes (as cited in Carnoy, Loeb & Smith, 2001, p. 1). According to Goertz, Duffy, and Le Floch, “Forty-eight states use a state assessment as the principal indicator of school performance. The other two states – Iowa and Nebraska – require their districts to test students in specified grades or grade spans, but leave the choice of assessment instrument to the locality. States differ, however, in the subjects and grades they assess and the types of tests they administer” (2001, p. 3). Accountability and its foundation, large scale state standards-based assessments, were born out of a relatively simple logic. As Richard Elmore pointed out (2004),

Schools and school systems should be held accountable for their contributions to student learning. Society should communicate its expectations for what students should know and be able to do in the form of standards; both for what should be taught and for what
students should be able to demonstrate about their learning. School administrators and policy makers, at the state, district, and school level, should regularly evaluate whether teachers are teaching what they are expected to teach and whether students can demonstrate what they are expected to learn. The fundamental unit of accountability should be the school, because that is the organizational unit where teaching and learning actually occurs. Evidence from evaluations of teaching and student performance should be used to improve teaching and learning, and ultimately, to allocate rewards and sanctions (Elmore, Abelman et al., 1996, as cited in Elmore, 2004, p. 4).

Based on this reasoning, Elmore contended that high stakes standards-based testing, conflicted with the traditional institutional structure of our schools and was unlikely to affect genuine school improvement until student results were tied to individual teachers. Elmore (2004) defined the institutional structure of schools as a “loose coupling” meaning, “The technical core of education… resides in individual classrooms, not in the organizations that surround them” (pp. 5, 6). The management of schools then is essentially the management of the structures and processes around instruction, but not the management of instruction. As a result, schools aim to:

1) Protect teachers from outside intrusions in their highly uncertain and murky work, and

2) Create an appearance of rational management of the technical core, so as to allay the uncertainties of the public about the actual quality or legitimacy of what is happening in the technical core” (Elmore, 2004, p. 6).

The product of this conflict between high stakes standards-based testing and the loose coupling of our public schools is a flawed accountability system. For instance, high stakes testing directly influences the instructional core of schools, making classroom instruction an issue of public
policy. Also, standards-based accountability forces schools to monitor, by student subgroup populations, why some students attain proficiency in certain subject areas and others do not. Lastly, Elmore asserted that standards-based testing weakens local control of schools because it focuses mainly on individual schools and not school districts as the principal unit of accountability. In spite of this disagreement, school and district accountability tied to state test scores is not likely to go away if for no other reason than that state legislators are saddled with the responsibility of accounting for public school expenditures (Elmore, 2004).

More recent changes to educational accountability policy under Race to the Top (RTT), a $4.35 billion competition sponsored by the U.S. Department of Education through the American Recovery and Reinvestment Act of 2009, raised the school reform stakes even higher by making principals directly answerable for the success or failure of their teachers and schools (Karhuse, 2012). The RTT grant application encouraged states to implement education policies in six core areas, and included the criteria by which states’ plans were evaluated. The criteria for the teacher evaluation area included whether states proposed to:

1) Establish clear approaches to measuring student achievement growth for individual students.

2) Design and implement rigorous, transparent, and fair evaluation systems for teachers.

3) Differentiate effectiveness using multiple rating categories that take student achievement growth into account as a significant factor and are designed with teacher involvement.

4) Conduct annual evaluations that include timely and constructive feedback and provide teachers with data on student achievement growth for their students, classes, and schools.
5) Use evaluations to inform decisions about staff development, compensation, promotion, tenure, certification, and removal of ineffective teachers (Hallgren, James-Burduny, & Perez-Johnson, 2014).

Research has shown some of the teacher evaluation policies promoted by RTT may produce more valid and reliable estimates of teacher quality. For instance, Mihaly, McCaffrey, Staiger, and Lockwood (2013) and Kane, Taylor, Tyler, and Wooten (2010) said evaluations using more than one measure more reliably capture teacher performance, and Weisberg, et al. (2009) and Gordon, Kane, and Staiger (2006) advocated using more than two rating categories better differentiates teacher performance. The goal of policies like RTT was to produce better information to improve teacher performance and therefore student achievement.

Teacher Effectiveness

While we know good teaching is important, it’s far less clear what makes for a good or bad teacher. Research dating back to the 1966 release of Equality of Educational Opportunity, known as the Coleman Report, indicated school-based factors such as principals, class sizes, facilities, curriculum, technology, teacher quality quantified by years of experience, education level measured by degrees obtained, and teachers’ performance on a vocabulary test had considerably less influence on student achievement than non-school factors, such as, students’ socioeconomic background (so-called peer effects). More recent research has confirmed that school factors, which principals have some control over, like high-quality teachers, do in fact raise student performance (Goldhaber, 2002). Winfield (1986) and Page (1991) identified teacher motivation and expectations for students as crucial factors for success with students. McLaughlin and Talbert (1993) and Knapp (1995) said teachers who assumed personal and professional
responsibility for students’ success and motivation, and maintained high standards were more likely to have students who succeeded. In contrast, teachers who believed students were responsible for their own motivation toward school, or some students were not capable of adequate performance were more likely to have students who were less successful. Researchers studying Tennessee data linking teachers to student achievement scores found the effectiveness of teachers had a more “value added” influence on student achievement than any other school factor with average student gains ranging from 14 to 52 percentile points a year depending on teacher quality, with these effects persisting for years after a student had a particular teacher (Sanders & Rivers, 1996). Measurable and immeasurable school characteristics were found to account for approximately 21 percent of the variation in student achievement, while about 60 percent of differences in student test scores could be explained by non-school individual and family background characteristics. Teachers’ subject-matter knowledge as measured by education levels (degrees) and experience levels (years), the easiest to measure teacher attributes and the sole determinants of teachers’ salaries, were not found to consistently, or positively influence student learning. Ehrenberg and Brewer (1994) found students scored higher on standardized exams if their teachers attended more selective undergraduate institutions, even though the effect of a degree often hinged on labor market conditions at the time the teacher was hired and whether the degree was specific to the subject being taught. There was also little evidence that experience beyond the first couple of years in the classroom made one a better teacher. How teachers teach, their pedagogical knowledge, measured by their performance on licensure exams tended to have strong positive correlations with the average student performance on standardized tests, however, it was not clear if this reflected the influence of subject-matter knowledge or pedagogical training since most state licensure exams tested both. Studies by
Greenwald et al. (1996) and Ferguson and Ladd (1996) found that teachers’ academic skills, such as aggregate teacher scores on the SAT or ACT and tests of verbal ability, showed a positive relationship to aggregate student performance on standardized tests. However, aggregate (school or school district) studies were unclear about whether higher-scoring teachers led to higher-scoring students or whether affluent districts which tended to have higher-achieving students also tended to hire higher-scoring teachers (Goldhaber, 2002). Teacher quality studies varied in value, looked at different types of students, subjects, and grade levels, and randomized experiments with detailed controls for student background characteristics (such as previous academic achievement) were rare in education research. Furthermore, referencing Hanushek et al. (1998), Rockoff (2003), Goldhaber, et al. (1999), Rowan et al. (2002), and Nye et al. (2004) Di Carlo (2010) posted this on the Shanker Blog website, “Observable and unobservable schooling factors explain roughly 20 percent, most of this (10-15 percent) being teacher effects. The rest of the variation (about 20 percent) is unexplained (error). In other words, though precise estimates vary, the preponderance of evidence shows that achievement differences between students are overwhelmingly attributable to factors outside of schools and classrooms” (2010, para. 6). This is certainly not to say that teachers don’t matter; of course they do. But some scholars have contended we have to develop nuanced explanations that take into account the complexities and the unquantifiability of the many moving parts of public schools.

Goldhaber (2002) confirmed principals have historically used the most easily quantifiable supply-side measures of teacher quality to screen and hire candidates: (a) certification, (b) education level, and (c) experience even though there is little empirical evidence these metrics are associated with higher student achievement. Studies on teacher certification have been largely inconclusive (partially due to no nationwide standard system of licensure), teacher
education levels made a difference only when the advanced degree was related to the subject taught, and the benefits of additional years of experience leveled off after the first few years of a teacher’s career. Ehrenberg and Brewer (1995) specified academic skills such as a teacher’s verbal ability were more likely to accurately predict effectiveness, but still the available evidence was mixed (2002). The intention of state teacher evaluation legislation like SB 191 was to increase teacher accountability and thereby teacher quality, not the other way around. The national dialogue has been more about evaluating and grading teachers, and less about coaching and nurturing them. How principals apply teacher evaluation policies and instruments is the key leverage point to improving the overall quality of public schools, but has been rarely talked about in education research. Although economics literature discusses employee compensation exhaustively, comparatively few empirical studies examined the factors that employers considered when hiring, promoting, or dismissing workers. The notable exception was the large body of work devoted to determining whether, and under what circumstances, employers discriminated against racial minorities and women. Yet discrimination research ironically has not provided solid evidence that employers considered factors related to effectiveness and productivity when hiring or dismissing employees, only whether or not they were discriminatory.

The large-scale, multi-year Measures of Effective Teaching (MET) project, funded by the Bill and Melinda Gates Foundation in 2013, was a collaboration between 3,000 teacher volunteers, administrators, union leaders, and education researchers (from Dartmouth College, Harvard University, the University of Michigan, the University of Virginia and Stanford University) in six school districts (Charlotte-Mecklenburg, NC Schools; Dallas, TX Independent School District; Denver, CO Public Schools; Hillsborough County, FL Public Schools; Memphis, TN City Schools; and New York City, NY Department of Education) that investigated
if effective teaching could be identified and developed. The MET project used the following multiple measures of data collection:

1) The Tripod Student Perception Survey (TSPS) developed by Harvard’s Ron Ferguson to provide teachers feedback on student perceptions of their classroom experiences;

2) The Content Knowledge for Teaching (CKT) test developed by the Educational Testing Service (ETS) to measure teacher understanding of instructional strategies using written assessment; and

3) These classroom observation protocols to help school principals evaluate teaching effectiveness:
   - The Framework for Teaching by Charlotte Danielson;
   - The Classroom Assessment Scoring System (CLASS) from the University of Virginia;
   - The Mathematical Quality of Instruction (MQI) from the University of Michigan;
   - The Protocol for Language Arts Teaching Observations (PLATO) from Stanford University;
   - The Quality of Science Teaching (QST) also from Stanford University; and
   - The UTeach Teacher Observation Protocol (UTOP) created by the University of Texas-Austin to assess math and science instruction.

The study randomly assigned students, looked at professional practice, and measured student achievement using each school district’s state standardized assessment and these supplemental tests: Stanford 9 Open-Ended Reading Assessment grades 4-8; Balanced Assessment of Mathematics (BAM) grades 4-8, and ACT QualityCare series for Algebra I, English 9, and
Effective teaching can be measured;
2) Balanced weights indicate multiple aspects of effective teaching;
3) Adding a second observer increases reliability significantly more than having the
same observer score an additional lesson;
   a) Additional shorter observations can increase reliability;
   b) Although school administrators rate their own teachers somewhat higher than do
outside observers, how they rank their teachers’ practice is very similar and
teachers’ own administrators actually discern bigger differences in teaching
practice, which increases reliability; and
   c) Adding observations by observers from outside a teacher’s school to those carried
out by a teacher’s own administrator can provide an ongoing check against in-
school bias (Measures of Effective Teaching Policy and Practice Brief Ensuring
Fair and Reliable Measures of Effective Teaching: Culminating Findings from the
MET Project’s Three-Year Study, 2013b, pp. 4-5).

Additionally, to enhance teacher effectiveness Leithwood and Montgomery (1982) and
Wahlstrom and Louis (2008) specified the design of an evaluation system should also link
student learning measures to teachers’ professional development plan.

The Principal’s Role in Effective Teacher Evaluation

While it is difficult to reduce good teaching down to an isolated observation record from
an evaluation rubric or checklist, informative classroom observation conducted regularly was key
to improving professional practice in schools. The effectiveness of any instructional strategy can only be determined by evidence of its effect on student learning (DuFour & Marzano, 2011). The most influential strategy for improving teaching and learning was not micromanaging instruction, but creating the collaborative culture and collective responsibility of a professional learning community (PLC). Research conducted by the Center on Organization and Restructuring of Schools (Newmann & Wehlage, 1995), the National Commission on Teaching and America’s Future (Carroll, Fulton, & Doerr, 2010), the Annenberg Institute for School Reform (2005), the Wallace Foundation (Louis, Leithwood, Wahlstrom, & Anderson, 2010), and the American Educational Research Association (Holland, 2005) confirmed the positive effect of PLCs on adult and student learning (DuFour & Mattos, 2013). In a collaborative school culture grounded in reciprocal accountability, teachers are answerable to results, and publicly recognize and celebrate their progress. Acknowledgement and support from the principal was essential to sustaining continuous improvement (Heath & Heath, 2010).

Means, Padilla and Gallagher (2010) believed principals needed to provide teachers with timely, standardized feedback on student learning progress to build a culture of collective responsibility for learner-focused outcomes, especially considering the prevalence of benchmark assessment systems in public schools. Formative feedback in education was typically studied in terms of student learning (Brophy, 1981; Schwartz & White, 2000; Black & William, 1998). DeNisi and Kluger (2000) suggested that to motivate and improve performance feedback should focus on the task and task performance, not on the individual person or the person’s self-concept. Spillane, Halverson and Diamond (2004) said leadership was composed of macro- and micro-tasks. Macro-tasks referred to general organizational tasks, such as providing adequate resources, planning, and designing professional development. Micro-tasks broke these general
responsibilities down into school principals’ day-to-day activities. Through these substantive and symbolic interactions with the school community, principals could be viewed as instructional leaders. Knapp, et al. said evaluations of teacher performance should be based on multiple data sources including a self-evaluation (2003). Principals must dedicate time to formative and summative teacher evaluation with regular feedback (High & Achilles, 1986; Marzano, et al., 2005), and find resources for teachers to improve practice consistent with school wide improvement planning (Nelson & Sassi, 2005; Wahlstrom & Louis, 2008).

Overall, evaluations are intended to measure a faculty’s ability to deliver the school’s instructional initiatives (Hallinger & Heck, 1996). The evaluation process should support numerous classroom visits by multiple observers (Blase & Blase, 1999), document poor teaching, and provide useful feedback to accomplished teachers (Quint et al., 2007). Effective principals continuously seek outside expertise, cultivate a “critical friends” group to get perspective on school progress, and maintain strong relations with the district to influence priorities (Halverson & Thomas, 2007; Kelley & Shaw, 2009).

**Sensemaking**

Sensemaking is the process by which people assign meaning to experience. Weick (1995) said it’s “the making of sense” that happens when principals notice, observe, collect, and organize information in relation to something in their environment and what they already know (p. 4). Starbuck and Milliken (1988) indicated sensemaking places stimuli, or data, into a structure that enable people, or groups, “to comprehend, understand, explain, attribute, extrapolate, and predict” (in Ancona, 2012, p. 4). Blumer (1969) explained symbolic interactionism, the prelude to sensemaking theory, this way:
The actor selects, checks, suspends, and transforms the meanings in light of the situation in which he is placed and the direction of his action… interpretation should not be regarded as a mere automatic application of established meanings but as a formative process in which meanings are used and revised as instruments for the guidance and formation of action (p. 5).

Weick, Sutcliffe, and Obstfeld, (2005) wrote interpretation was just one part of sensemaking. A person did not simply interpret and then act. “Sensemaking is the activity that enables us to turn the ongoing complexity of the world into ‘a situation that is comprehended explicitly and serves as a springboard into action’” (p. 409 in Acona, 2012, p. 4). Information and knowledge was first noticed, understood in relation to some situation or context, and then used for decision-making. Each of step depended upon the resources, cognitions, and capabilities of principals working in schools (Spillane & Miele, 2007). Moreover, social interactions could actually define sensemaking. What individuals attended to and interpreted wasn’t done in isolation; it was intrinsically connected to the situational factors that influenced behavior (Spillane & Miele, 2007). Separating personal beliefs about the way the world worked from situational factors like rules and routines that shaped what a principal was required or permitted to do was difficult because sensemaking needed social interactions. Blumer (1969) suggested,

> Individuals act toward others, institutions and objects based on the personal meaning of things, the meaning of these things derives from social interactions and these meanings are modified through an interpretative process used by individuals encountering and engaging these other people, institutions and objects (p. 2).

Sensemaking thereby involved a “crucial set of elements, including self, action, interaction, interpretation, meaning and joint action” (Weick, 1995, p. 41).
While symbolic interactionism was foundational to understanding how information was used by principals to evaluate teachers, researchers have not overcome the challenge of identifying under what conditions sensemaking processes actually transformed schema, mental models, or practices. To date, this limitation has resulted in us knowing very little about what types of information and what conditions influenced principals’ interpretations and decision making. In addition, we lack empirical evidence and theories to guide investigations seeking to understand new forms of data like the kind SB 191 has made available. Without better understanding what principals noticed and interpreted, and under what conditions, it was difficult to identify and predict how educator evaluation legislation could transform practice and lead to improved outcomes. Similarly, it was difficult to estimate how the implementation of the new 5D+ evaluation system was influenced by the context of PVSD (Coburn & Turner, 2012b).

**Principal Decision-making in Teacher Performance**

Do principals believe their evaluations of teachers’ professional practice improve student performance? Especially when, as Goldhaber (2002) asserted, the importance of teacher quality on student learning is widely accepted in spite of the little direct empirical evidence available. Teachers have a profound effect, both positive and negative, and principals make a permanent, long-term financial investment when they choose to hire and retain teachers in the existing tenure system. A principal essentially makes a 1.62 million dollar vote of confidence with public school funds from local tax payers when he/she tenures a teacher with a fourth consecutive year after three years of satisfactory service (assuming the teacher teaches for 27 more years multiplied by $60,000 in salary and benefits equals 1.62 million). For these reasons principal efficacy in evaluating teachers is paramount, yet the public perception is few principals exercise their
authority to non-renew teachers, which lead some to question principal effectiveness. “In 1986, W. Edwards Deming argued that leaders must ‘drive out fear’ from their organizations because appeals to fear resulted in short-term thinking, fostered competition rather than collaboration, and served as a barrier to continual improvement” (DuFour & Mattos, 2013, n.p.).

In 2004 the Chicago Public Schools (CPS) and the Chicago Teachers Union (CTU) signed a new collective bargaining agreement that changed policy and in effect gave principals the flexibility to dismiss probationary teachers for any reason without the documentation and hearing process previously required. Matching information on CPS teachers who were eligible for dismissal with teachers who were actually dismissed Jacob (2011) estimated the comparative weight school principals placed on a variety of teacher characteristics, including productivity, teacher absences, value-added measures, and demographic characteristics in deciding which teachers to dismiss. Jacob found principals were significantly more likely to dismiss teachers who are frequently absent and received poor evaluations in the past, and were significantly less likely to dismiss teachers with stronger educational qualifications as measured by the competitiveness of their undergraduate college, whether they ever failed the teacher certification exam, and had a Master’s degree. Furthermore, elementary teachers who were dismissed had significantly lower value-added with regard to student achievement in prior years compared with their peers who were not dismissed, and dismissed teachers who were subsequently re-hired by a different school were substantially more likely to be dismissed again relative to first-year teachers in that school.

Jacob’s analysis also revealed certain teacher demographic characteristics can be associated with the probability of dismissal, i.e. principals are more likely to dismiss male teachers, and older teachers, especially if older teachers work in buildings with younger
principals. That being said, it is worth noting that many principals, including those in some of the worst performing Chicago schools, did not dismiss any teachers at all despite how easy it was under the new policy. This appeared to be consistent with findings from previous studies that noted existing teacher contracts in many school districts actually offered more flexibility than was commonly believed yet administrators rarely took advantage of such flexibility (Ballou, 2000; Hess & Loup, 2008; Price 2009). This suggested that issues such as teacher supply, union influence, and/or social norms governing employment relations had a bigger effect on principal decision-making than state legislators realized (pp. 429-430). Also, 50.6% of first-year probationary teachers and over 60% of third-and fourth-year probationary teachers dismissed in spring of 2005 (whether due to budget-related staffing cuts or poor performance) were rehired by a CPS school in fall 2005 (p. 414). Furthermore, relative to principal preferences for worker characteristics, employers in general tended to prioritize “soft skills” like motivation, enthusiasm, customer interaction, and communication skills that are generally unobservable to an econometrician (Murnane & Levy, 1996; Moss & Tilly, 1995). This was consistent with the research of Harris, Rutledge, Ingle, and Thompson (2006), which found that principals favor teachers who demonstrate strong teaching skills, are caring, experienced, enthusiastic, knowledgeable of their subject matter, have the ability to work well with others, and communicate effectively.

Mager and Pipe (1997) proposed the basic purpose and underlying motivation for solving virtually any performance problems should be identifying the mismatch or difference between what is and what should be. When talking about human behavior, they referred to this performance discrepancy as the difference between what people actually are doing, or not doing, and what they should be doing, or not doing. Senge (1990) talked about this concept as the
creative tension between current reality and perceived reality. If teacher performance was not where it needed to be, principals were responsible for providing the information, coaching, and training necessary to address areas of improvement.

Helping a teacher close that gap and build a reality that matches the vision of high level performance requires principals to understand the difference between what social scientists refer to as a heuristic task and an algorithmic task. Pink (2009) put it this way, “An algorithmic task is one in which you follow a set of established instructions down a single pathway to one conclusion... a heuristic task is the opposite. Precisely because no algorithm exists for it, you have to experiment with possibilities and devise novel solutions” (p. 111). Schlechty (2011) offered that teaching is viewed as an algorithmic task based on 19th and 20th century thinking regarding human motivation, but in actuality it’s a heuristic task. He said,

The language most teachers have been taught to use to talk about their task is that of bureaucrats and planners rather than the language of design. They are taught to speak of goals, objectives, and time lines rather than of motives, needs, and values. They speak of measurement more than prototypes and data more than ideas. The language they use is often instrumental and cold as opposed to expressive and inviting. Students are products rather than neophytes to be inducted into the ways of knowledge work, and teachers are professional employees rather than leaders of children and of parents and communities” (p. 50).

Accordingly, principals are challenged with completing the heuristic task of improving teacher performance within an increasingly algorithmic evaluation enterprise controlled by legislated rules, codified procedures, pre-determined goals, and quantifiable results. The limited research on principal decision-making regarding teacher performance leaves many unanswered
questions: Do principals consider the socio-cultural political context of evaluation decisions? Is the internal and/or external culture of a school and/or district supportive of teacher non-renewals and dismissals? Do principals feel supported by their district-level supervisors? Will the decision or process negatively impact staff morale, or incite the teachers’ union to criticize the principal? Is the available pool of qualified applicants likely to yield a better teacher? Do principals have confidence in their ability to identify effective professional practice and write legally defensible evaluations? Do principals have the time to adequately complete and document classroom observations? Is evaluation a viable way to improve a school?

**Confronting Poor Teaching**

Personnel decisions should ensure a highly qualified and competent faculty. Principals need the autonomy and discretion to hire and assign teachers (Halverson & Thomas, 2007; Kelley & Shaw, 2009). Effective principals must be willing to confront teachers who fail to honor their obligations to their students and their commitments to their colleagues. It is the responsibility of the principal to communicate that an individual teacher cannot disregard the district-adopted curriculum, refuse to administer common assessments, treat students or staff with disrespect, or opt out of collaborative team processes and meetings. Bryk, Sebring, Allensworth, Luppescu, and Easton (2010) emphasized principals need to be willing to use their authority to break down the walls of educator isolation and create norms that foster the moral imperative and professional responsibility of collective responsibility for student learning.

According to education title 22, article 63-301 teacher employment, compensation, and dismissal part 3 of the Colorado Revised Statutes (C.R.S.) grounds for dismissal,
A teacher may be dismissed for physical or mental disability, incompetency, neglect of duty, immorality, unsatisfactory performance, insubordination, the conviction of a felony or the acceptance of a guilty plea, a plea of nolo contendere, or a deferred sentence for a felony, or other good and just cause. No teacher shall be dismissed for temporary illness, leave of absence previously approved by the board, or military leave of absence pursuant to article 3 of title 28, C.R.S. (C.R.S. 22-63-301, 2012).

The procedure for dismissal, judicial review, is explained in the subsequent C.R.S. 22-63-302. Regarding unsatisfactory performance, only when a teacher’s low value added scores agreed with the principal’s observational assessment of that teacher’s performance was the teacher likely to be dismissed, which is not much different, and maybe worse, than the previous system. While discussing two relevant articles, *Managing the Teacher Workforce* by Goldhaber and Theobald and *Principled Principals* by Jacob, editor Eric Hanushek stated in an Education Next podcast with author Martin West, “Very few studies have looked systematically at the issue of teacher dismissal before these, in part because there have been very few evaluations that have been meaningful, and very few related dismissals aimed at producing a higher quality teaching force” (2011). This suggested that despite the promise of SB 191 to make dismissing teachers more scientific through an expedited procedure, removing underperforming teachers remained a political and social process that needed to account for the social forces that upheld the status quo.

In almost all public school districts teacher layoffs and dismissals were highly regulated and contractual. In situations where teacher positions were eliminated due to enrollment changes or a budget deficit known as a reduction-in-force (RIF), the seniority clause of the collective bargaining agreement took affect and demanded the least experienced teachers got let go first, commonly known as last-in-first-out (LIFO). Historically, was difficult for principals to dismiss
teachers outside the auspices of a RIF, as collective bargaining agreements provided considerable protection for tenured teachers that made it time-prohibitive for principals to dismiss teachers for cause (Jacob, 2011). As a point of fact, most teacher mobility studies looked predominantly at the voluntary separation of teachers from their positions, i.e. they chose to leave to go to higher paying districts, with a more affluent or less at-risk student population, better facilities, and improved test scores. Yet unlike teacher mobility studies, this study considered involuntary separations as an indicator of a principal’s level of conviction, or whether or not to dismiss a teacher for performance, as they were the most difficult summative evaluation decisions principals made.

Prior to the 4.35 billion in Race to the Top incentives initiated by the Obama administration in July 2009, the American Federation of Teachers (AFT) and the National Education Association (NEA) communicated their position on the constitutional and statutory protections against unjust dismissal in Peer Assistance and Peer Review: An AFT/NEA Handbook. The handbook asserted the U.S. Supreme Court held that once a teacher has completed his/her probationary period under state law the teacher had a vested property right to current and future employment in his/her position under the due process clause of the fourteenth amendment and must be given reasons for his/her dismissal and an opportunity to be heard. The Fourteenth Amendment to the U.S. Constitution provided no state shall deprive any person of life, liberty or property, and public school districts are considered extensions of the state and tenure was a property interest. In the case of dismissal for cause, due process required adequate notice to the teacher and a fair hearing before an impartial decisionmaker and the opportunity to present evidence prior to termination (1998). However, contrary to common belief Elder stated (2004) it was, “Not impossible or even very difficult to weed out the deadwood… some
behaviors are so egregious that dismissal is justified by the act alone” (p. 26). More specifically, Elder identified “five easy calls” for firing: “1) Lateness, 2) Failure or refusal to report child abuse, 3) Violation of confidentiality, 4) Sexual harassment in the workplace, and 5) Ethnic slurs, lying, stealing, cheating, and illegal acts” (2004, pp. 26-30). To further explore if and how principals responded to ineffective teaching and the perceived likelihood they could hire a better teacher than the one terminated, we turned our attention to what principals paid attention to and were influenced by as expert observers of teaching and learning.
Chapter 3: Research Methodology

Colorado Senate Bill 10-191, the Great Teachers and Leaders Bill (SB 191) passed in 2010 in support of the U.S. Department of Education’s Race to the Top program and the No Child Left Behind Act, mandated increased teacher and principal accountability by affirming improved teacher evaluation would advance K-12 schooling. As a result, the 27,000-student Pioneer Valley School District (PVSD) developed and implemented a new evaluation protocol. The new process weighted 50% of teacher and principal summative performance ratings on professional practice using the 5 Dimensions+ Teacher Evaluation Rubric (5D+), and 50% on student achievement data using a PVSD-created three-tiered value-added model (VAM).

Because school principals are responsible for evaluating teachers, this research used Ajzen’s Theory of Planned Behavior (2006) as the conceptual framework to simultaneously consider how attitude toward a behavior (behavioral beliefs), social pressure (normative beliefs), and the level of behavioral control (control beliefs) influenced principals’ decisions when assessing teacher performance in an environment of conflicting social forces and the collegial interactions which may contribute to or challenge relational trust and maintenance of the status quo. While the content of teacher evaluations written by building principals was typically unavailable to outsiders due to the confidential nature of personnel files, the outcomes of evaluation decisions were keenly felt by school and district insiders, influenced by the broader educational environment, and were inherently multilevel. The purpose of this study then was to explore the factors that influenced principal behavior in assessing teacher performance. Two major research questions were investigated:

1) What factors do principals perceive as influencing and shaping how they evaluate teachers’ professional practice?
2) How do principals perceive the manner in which these factors influence and guide their performance ratings of teachers? In what ways did the principals feel these factors influenced their performance ratings of teachers?

In-depth one-on-one participant interviews were the primary means of data collection and analyzed along with 5D+ evaluation documents in this exploratory case study to develop an understanding of principals’ perceptions of the main factors, and the ways in which those factors influenced their decision making when evaluating of teachers’ professional practice.

**Justification for a Qualitative Approach**

The research problem discussed in Chapter One required an approach that was field-based, descriptive, and holistic. Consequently, this study employed a qualitative design consistent with the Interpretivist paradigm to understand, “What’s going on here?” (Rist, 1982, p. 440). Qualitative research seeks to understand the meaning people have constructed from their experiences in a particular social context and their interactions with others in that context (Merriam, 1998). As Patton (1990) described:

> Qualitative research is an effort to understand situations in their uniqueness as part of a particular context and the interactions there. This understanding is an end in itself, so that it is not attempting to predict what may happen in the future necessarily, but to understand the nature of that setting – what it means for participants to be in that setting, what their lives are like, what is going on for them, what their meanings are, what the world looks like in that particular setting (p.1).
This study built from a description of individual principal’s beliefs and actions to provide a greater understanding of the personal, social, political, technical, and cultural forces that influenced principals in their teacher evaluation role.

Qualitative methodology also fit well with answering the research questions to accomplish the goal of this study – identify and understand those factors that principal perceive as influencing their decision making in evaluating teachers’ professional practice. The assumption was principal observations, experiences, and decisions were nested in school settings, that were embedded in district systems, that were also part of the broader educational policy environment. Therefore, individual realities were inherently multilevel, and varied across school settings (Talbert and McLaughlin, 1999). As Talbert and McLaughlin pointed out,

The theoretical lens of rational organization theory describes these contexts in terms of educational policies, resource levels, and allocation patterns, and program and governance structures. Although the structural hierarchy of the school administrative system presents a nested model of school organization, the policy system does not operate in a linear and hierarchical manner. In other words, lower level policies are not predicated on high-level policies; furthermore, higher level policies can have effects on practice that are not mediated by middle-level policies of the system (1999, p. 205).

A qualitative approach was conducive to this task and allowed the researcher to see the world as the principals see it. The phenomenon studied was embedded in and inseparable from the context of a particular school and its personnel. Truth emerged not as an objective view but as a composite picture of principals’ interpretation of reality. This resulted in the social, cultural, and historical understanding of the settings in which principals evaluated teachers. Because SB 191 was not yet fully implemented in Colorado when the study began, it was difficult to measure
the phenomenon directly. Nevertheless, illustrative examples of principal perceptions and practice were gathered through an interview process in which principals shared descriptive feedback about their unique and common experiences implementing the teacher evaluation process.

**Rationale for a Case Study**

The research design utilized to understand the multiple realities principals experienced when evaluating teachers was the case study. According to Creswell (2003), the case study approach allows a researcher to delve deeply into an event, activity, or process, of one or more individuals, using a variety of data collection procedures, within a particular context bounded by time. Yin (2009) proposed at least five different applications for case study research, with the most typical being to explain the hypothesized causal links of phenomena that are too complex for experimental designs. Another application of the case study is to describe a phenomena and the real life context in which it occurs. The third application is to illustrate topics when evaluating a particular social phenomenon. Fourth, a case study can be used to explore situations in which the phenomena being evaluated have no distinct set of outcomes. The fifth and final application is that of metaevaluation, or the study of an evaluation study.

This research emphasized the explanatory and descriptive applications of case study design. It allowed for an examination of the interaction between context and principals’ beliefs, understandings, and assumptions to be explored.

**Site Selection and Principal Recruitment**
This study used a convenient, purposeful sample by predetermining the Pioneer Valley School District (PVSD) as the site and fifteen PVSD principals as the participants. As a current middle school principal in the PVSD, the researcher had easy access to the participants, some knowledge of their experiences, and familiarity with the teacher evaluation tool being used (see Appendix F). The investigator invited all fifteen principals, personally or over the phone, to take part in this study and sent a follow up email invitation separately to each participant in October of 2014 (see Appendix A). Furthermore, upon agreement to participate, the researcher contacted each principal in person, by telephone, and email as necessary to explain the objectives of the case study, the nature and expectations of the research, the amount of time the interviews would take, and the information to be shared. Relationships with the principal participants were nurtured so they felt personally involved in the outcomes of the study. It also allowed the researcher to check that participants had pertinent information to share, were willing to include the interview in their schedule of competing tasks, and felt the results would include information they valued and that informed their work roles and responsibilities. One-third of PVSD principals were interviewed to gather robust information about the perceptions they held with regard to the new teacher performance evaluation system. This purposeful sample emphasized representativeness, as the school principals chosen embodied the diversity of schools across PVSD which varied in size, level, resources, and socio-economic status of the families served.

The Pioneer Valley School District (PVSD) is a high performance and high growth school district in northern Colorado that employs 3,476 people (106 administrators, 1,903 licensed, 1,467 and non-licensed) and educates 27,510 students (74% White, 17% Hispanic, 3% Asian, 1% Black). In addition, 31% of students qualify for free or reduced lunch support, 10% are identified as gifted and talented, 9% are English Language Learners, and 8% receive special
education services. The district’s 1,856 square miles cover an area larger than the state of Rhode Island along Colorado’s Front Range. PVSD’s open enrollment (school choice) program allows families to select the school that best meets their child’s educational needs. Nearly 70% of PVSD families choose to send their children to the neighborhood school in which they live. Space permitting, families may choose to send their child to a neighborhood or choice school outside their attendance area if they provided their own transportation. PVSD transports 12,000 students daily on 117 propane powered school buses to full day kindergarten and neighborhood schools. PVSD is made up of 31 elementary schools, 10 middle schools, 7 high schools, 2 charter schools, and 1 K-12 online school. Students performed 8 - 15% higher than students around the state in all subjects and in all 27 areas tested on the 2012 Transitional Colorado Assessment Program (TCAP). The dropout rate was 1.3% and the graduation rate was 84% in 2012. PVSD spends roughly $7,750 per student and in November 2010 voters approved a 120 million dollar bond and 16 million dollar mill levy to restore previously cut positions, refresh technology, provide skills for workforce development and college readiness, and increase budget allocations to schools (2011-12 Accountability Report, Pioneer Valley School District).

Confidentiality and Anonymity

The Pennsylvania State University Office for Research Protections for Human Participants Research Institutional Review Board (IRB) approval of this project as exempt research according to its policies and applicable federal regulations validated the research methodology, confirmed the protection of human participants’ rights, and minimized known risks. Maintaining confidentiality and anonymity was an obligatory ethical component of this investigation. Participants’ names and identities were masked throughout the study, and are not
identifiable by the reader. However, principal participants could be recognized by other participants depending on their knowledge of identifying features of their statements, positions, or personal and social characteristics. Although this level of identification was unlikely, it was a necessary and important potential consideration of this research (Creswell, 2003).

All participants were provided an informed consent form to participate in a social science research study (see Appendix B), authorized by the Office for Research Protections, clarifying confidentiality protections. Strict confidentiality was maintained at all times through the use of pseudonyms, rather than participants’ names. Only the lead investigator had a record of each participant’s identity. After data collection and analysis was completed, the researcher referred to participants by pseudonym only. During the study, records were stored by the researcher on a secure server, password protected, with a backup external hard drive. All personal demographic data collected was accessible only to the researcher, thus protecting confidentiality and anonymity. No participant, school, or district names were used. All personal, school, and district identifying information was deleted at the conclusion of the study. However, it was important to provide some general information and context relative to the principals and the schools in which they worked to give credibility to the findings. This was communicated in the consent form. Reporting was both individual and aggregate, and a final copy of the dissertation will be given to each principal upon completion of the study.

**Role of the Researcher as a Participant Observer**

As the researcher in a participant observer study, I was the primary instrument of data collection, analysis, and interpretation for this qualitative case study (Denzin & Lincoln, 2003). I served as an *emic*, or insider, a role clearly known by all the participants (Punch, 1998). Being a
PVSD middle school principal for the past ten years allowed me to gain an understanding of the leadership styles of the principals, the inner workings of some of their schools and the district, and the role of the teacher evaluation in PVSD. While being an insider in the district was beneficial, it was also a limitation, as I brought my own biases, opinions, and judgments to the data collection and analysis. I recruited and interacted directly with all participants through the in-depth interviews, which elicited a more sympathetic orientation from principal respondents and enhanced my ability to see, hear, and experience their realities within their normal working environment as PVSD principals. I gathered data purposefully during the conversational interviews by observing the interviewee’s body language and affect in addition to recording his or her words. As this inquiry progressed it became clear that what principals perceived as influencing their evaluations of teachers was expressive of their deeper values and beliefs. This immersion as a participant observer allowed me to learn from my own experience in the setting and personally reflect on an emerging holistic analysis of PVSD principals’ perceptions (Marshall & Rossman, 1999). I included interpretive comments in the interview field notes based on my understanding of their work. This may have been affected by the role and relationships I already had with the participants and the way participants reacted to me as a consequence of that social role (Goetz & LeCompte, 1984). I worked to build a trusting environment but my purpose, data collection, was always clear. I presented these relevant aspects of my own biases, assumptions, experiences, and expectations as a participant observer so the reader is aware of the possible impact on this study (Greenbank, 2003).

Moreover, Guba and Lincoln (1994) defined a researcher’s paradigm as the “basic belief system or worldview that guides the investigator” (p. 105). I worked from the constructivist paradigm to critically understand, make choices about, and communicate the assumption that
there isn’t a “real” objective world out there; that reality is relative, constructed through human relationships, and situated within the historical moment and social context (ontology).

According to Blumer (1969) Interpretivist theory, or symbolic interactionism, is rooted in three basic ideas. “The first premise is that human beings act toward things on the basis of the meanings that the things have for them… The second premise is that the meaning of such things is derived from, or arises out of, the social interaction that one has with one’s fellows. The third premise is that these meanings are handled in, and modified through an interpretive process used by the person in dealing with things he encounters” (p. 2). The Interpretivist paradigm is based on the premise that there is no objective reality, and the researcher is viewed as an investigator developing relationships while gathering information in an attempt to provide an organic picture of the natural setting (personal communications and classroom notes). Epistemologically, the researcher actively co-constructs knowledge with study participants as the “knower” could not be objective and not affect the outcome. Consequently, multiple meanings developed from the same data. Lastly, in disseminating the results I employed inductive reasoning and focused on empirical and logical narrative description (Creswell, 2009).

**Researcher identity.** Furthermore, who I am, my core values, and what I came into this project knowing and believing influenced how I engaged in the research process. As the sole instrument of data collection and analysis with a familiarity with the participant principals and a history with school district processes, I brought some preconceptions to the mechanics of this investigation. I managed this incredibly complicated interface with the participants intentionally to ensure the information contained in this study was what the participants said and not what the researcher thought. My intention was to promote the integrity of the process, and not appear self-indulgent.
I was born and raised on what was an upstate New York dairy farm that had been in my Irish-catholic family for many generations. By the time my parents were ready to enter the workforce, farming in New York’s Hudson Valley was becoming increasingly cost prohibitive. After graduating from college they opted for careers in public education. My father taught and coached high school football and rowing for five years before serving as an elementary and district administrator for twenty-six years. Upon retiring from public education he coordinated the teacher education program at a local private college for another twenty years. My mother spent thirty-one years as a second through sixth grade teacher, and for eight of those years served as the local teachers’ union chief building representative for her middle school. She then spent the first twelve years of her retirement supervising student teachers; in the same teacher education program my father oversaw. Growing up as the product of educators I thought everyone was supposed to find a career and stick it for thirty years. I was taught to respect my elders and not challenge authority. Children were meant to be seen and not heard. You did what you’re told, and saluted with both hands. The assumption was those who were in positions of leadership were responsible, ethical, and arrived there through their own merit.

My perspective as an adult and educator has evolved from my childhood, but my core values have not. I believe a free and appropriate public education is a civil right intended among other things to eradicate the effects of poverty, and if delivered effectively promotes both excellence and equity for students and families. These two goals do not necessarily have to be mutually exclusive. The three national awards my school has received over the past three years (2014 CAMLE School to Watch, 2014 USDE Green Ribbon School, and 2012 MetLife/NASSP Breakthrough School) and my experience turning a school slated for closure within three years of my hire date to the school with a wait list that receives more school choice applications than any
neighborhood secondary school in our district, in spite of its downtown location and fifty-five year old facility, has also colored my perspective. I’ve realized not all principals share my beliefs, competitive passion, or work ethic. The problem statement grew out of our elected representatives’ efforts to reform education by passing laws that many educators feared were an assault on tenure and rattled the foundation of public education. There appeared to be a mounting distrust for school teachers and principals. The theory was state legislators would not have mandated accountability laws like SB 191 had school administrators and union leaders been able to successfully “fix” public education over the past century. The research questions evolved from a desire to understand what other building principals thought about how this new legislation specifically and reform climate in general might affect them on the ground level in their day-to-day interactions with teachers. Since a review of the available research confirmed that less than one percent of teachers received an unsatisfactory summative rating annually across America, I wondered if principals really believed we could, or should be trying to “fire our way to effectiveness.” Conducting the data collection and analysis for this study proved to be a reflective exercise in balance as I worked to maintain objectivity when interfacing with the participants in spite of my strident advocacy for the historical ideals of public education with all of its flaws, as the cornerstone of our American democracy. In doing so, my intention was to create a meaningful study for educators and non-educators alike.

**Data Collection**

The research questions determined the data collection techniques and analysis (Howe & Eisenhart, 1990). The primary method of data collection was one-on-one in-depth interviews of fifteen PVSD principals. A main advantage to interviewing is it provided superior depth than any
other method of data collection (Cohen & Manion, 1994). One round of semi-structured interviews were administered individually; using four open-ended questions in a free response conversational format, with probing follow up questions, and were audio recorded. The interviews were scheduled over a two-week period according to the interviewees’ availability as suggested by Feagin, Orum, and Sjoberg (1991). Thirteen of the interviews occurred in the participants’ offices, and two were held in the researcher’s office. As Krathwohl (2009) stated, “Interviewing in the respondents’ home or office allows them to relax in their own territory, but phone calls and other business may create complications” (p. 301). As Rist (1982) believed, To conduct a good interview is to hold an interesting conversation. Ideas and perceptions are exchanged, information is shared, and participants come to know more about each other in the process. The importance in stressing the conversational aspect of interviewing is to reinforce the notion that qualitative work involves considerable human interaction (p. 443).

The fifteen interviews lasted from 55 minutes to 1 hour and 32 minutes of recorded time, and were completed in a two-week window of time between November 11 and November 24, 2014. The interview protocol and questions reflected the TPB and were designed to prompt principals’ opinions of their leadership practice and underlying theories of action as frontline implementers (see Appendix C). Follow up probing questions were asked for clarification. The belief-based, psychological, and internal constructs of the TPB model were addressed in the interview questions either directly, by asking participants about their overall attitude, or indirectly, by asking participants about specific behavioral beliefs. These direct and indirect questions generated different assumptions about participants’ underlying beliefs. In the interviews the principals’ reflected on how they exercised their professional responsibility and moral imperative
to accurately and honestly evaluate teachers using the 5D+ teacher evaluation rubric per the expectations outlined in the PVSD Employee Agreement (teacher contract). The individual nature of the interviews and the subsequent anonymity of principal responses eliminated any potential peer pressure between participants (Yin, 2009).

Principal respondents were encouraged to corroborate the rich primary source data they shared in their in-person interviews with evidence from other sources to provide further insight into their behavioral intentions, substantiate important findings, and offer a rationalization for potential differences in participant responses, but none did so. Prior to the interviews, the researcher gathered demographic information about the fifteen principal respondents and their schools, such as, gender, age, length of time at their current school, past performance reviews from their supervisors, Colorado Dept. of Ed. School Performance Framework (SPF) rating, enrollment, percent of students qualifying for free or reduced lunch support, etc.

**Data Analysis**

Data analysis occurred simultaneously with data collection. During the interviews the researcher continually assessed the field situation and made judgments on-site as necessary to adjust the line of analysis with participants in the most powerful and parsimonious way possible. As Rist (1982) described:

To state that data analysis occurs concurrent with data collection is only to acknowledge that fieldwork is not simply the mechanistic collection of predefined data from predefined sets of respondents. Rather, the entire time the researcher is in the field, there is a constant dialectic between collection and analysis (p. 445).
Rist (1982) identified seven analytic frameworks for the organization and presentation of qualitative data: (1) Role analysis, (2) Network analysis, (3) Natural history, (4) Thematic analysis, (5) Resource allocation, (6) Ritual and symbolism, and (7) Critical incidents. Based on the research questions and opportunities for data collection in this case study, thematic analysis was chosen to cluster and present the interview data gathered individually from participants in the form of written field notes and iTunes audio files (MPEG-4 .m4a). First, the interview audio files were transcribed and compared with the written field notes to identify key quotes and recurring pieces of text. Second, that descriptive text was used to identify themes, which originally were categorized around the TPB conceptual framework antecedents of attitude, subjective norms, and perceived behavioral control, which Ajzen (1991) identified govern our intentions and actions. When applied to what principals perceived as influencing their evaluation of teachers’ professional practice, the TPB suggested principals’ intentions to engage in crucial conversations with teachers about their performance were driven by their attitudes towards those interactions, perceived social pressure to do so, and their perceptions of control over the discussions. The more positive a principal felt about these three psychological constructs, the more likely the principal intended to and actually did evaluate teachers thoroughly and honestly. However, as the study evolved broader themes emerged and categories were expanded beyond the TPB factors to more accurately capture all the influences principals perceived as shaping how they evaluated teachers’ professional practice and the manner in which those factors guided their decisions. These categories provided insight into answering the research questions and are communicated in Chapter Four. As Owens (1982) pointed out,

An important element in the design of naturalistic research is starting with questions of broad scope and proceeding through a conceptual funnel – working with data all the
while, ever trying to more fully understand what the data mean – making decisions as to how to check and know to verify as the investigation unfolds (p. 11).

**Credibility and Trustworthiness**

McMillan and Schumacher defined credibility (2001) as, “The extent to which the results approximate reality and are judged to be trustworthy and reasonable” (p. 166). Moreover, trustworthiness refers to whether or not the researcher actually heard what he/she thought he/she heard and observed what he/she thought he/she observed. In other words, it’s the degree to which the participants’ and researcher’s perceptions and interpretations mean the same thing (McMillan & Schumacher, 2001). Merriam (1998) said ensuring research results are trustworthy involves conducting an investigation in an ethical manner. This case study, or bounded system (Smith, 1978), utilized procedures to increase the credibility and trustworthiness of the findings so the explanations of the phenomena matched the realities of the participants.

First, as the researcher and a participant observer, I went into this study with ten years of experience as a principal in the district. This provided me with an historical knowledge of the district, as I had already learned the language, and become “accepted, trusted, and ultimately relatively unnoticed” (Owens, 1982, p. 14). Relationships with fellow principals were nurtured to promote trust and open communication. As the interviews progressed, my perspective grew from initial impressions to a deeper understanding of the phenomenon (Owens, 1982). I made every effort to ask interview questions and follow up probes objectively to serve as in-the-moment member checks and build a chain of thick descriptive evidence so anyone interested in transferability had a thorough basis for comparison.
Second, patterns in the data were identified from detailed, concrete note taking as principals elaborated in the interviews. The researcher asked the participants during the interviews to clarify any abstract language so “low-inference descriptors” were documented as literally as possible (McMillan & Schumacher, 2001, p. 409). To guard against misinterpretation, key terms and ideas communicated by the participants were confirmed by the interview audio files, which provided a more complete record. Principals had little room to misinterpret the interview questions in this study. Directions were clear and the researcher’s contact information was provided to participants had questions arisen afterward. For these reasons and time limitations, participants were unable to read through their interview transcripts to change or alter any possible misconceptions.

Third, the researcher utilized member checks and peer consultation to ensure the researcher’s bias and perspective did not “color” the data. I confirmed observations and meanings with individual participants as necessary through casual follow up conversations in informal settings that occurred over the course of carrying out my day-to-day responsibilities as a PVSD principal. “Peer consultation provides opportunities while the inquiry is still in progress to check one’s thinking, to raise questions and concerns, and to talk through problems of which the investigator may or may not be aware” (Owens, 1982, p. 15). For the purposes of this study, I discussed the ongoing data collection and analysis with the dissertation chair, which assisted me in taking procedural and theoretical steps from concrete data to more developed concepts.

Overall, credibility and trustworthiness was enhanced by using a well-defined case study protocol to guide me in carrying out the data collection and analysis, and maintaining a chain of evidence or data trail so an external observer (the reader) is able to trace the steps from initial research questions to conclusions (Yin, 2009).
Limitations of a Case Study Design

Although the embedded case study was a useful methodology to answer the primary research questions posed in this study, it had shortcomings. One potential drawback was this research dealt with public school principals working in schools that served as learning institutions first and as research sites second, so principals could have declined the interview invitation and dropped out of the study, although none did. Also, the use of in-depth personal interviews as the primary data collection tool added to the possibility of researcher bias (Krathwohl, 1998), and the individual interviews did not allow for verbal discussion about conflicting perspectives amongst participants. Still another was the generalizability of the findings as this case study dealt with just one, relatively unique school district. When compared to the other 177 school districts in the state, PVSD could be viewed as a fairly privileged community with its healthy economy, local major university, high cost of real estate, and history of passing bonds and mill levies. Furthermore, its schools are well-maintained, annually outperform state averages in growth and achievement, offer a wide variety of arts and athletic programs, as well as advanced course offerings, and generally serve a more upper middle class and Anglo demographic.

This study was also limited by its sample size of fifteen PVSD principals (33%). If more time and money were available the study could have been expanded to include all of the principals in the school district. The non-random hiring of PVSD principals, short data collection time frame during November of 2014, and the principals’ familiarity with the researcher represented additional threats to the findings. Extreme opinions were another limitation, but given that this case study method did not require consensus to draw conclusions, outlying
opinions did not affect the outcome of the study. It’s also important to note that administrative tenure does not exist for public school principals in Colorado; they are at-will employees. This could have influenced how honestly principals answered the interview questions, had they feared being identified and perceived as insubordinate, or critical of the district. It’s also important to remember this was a purposeful sample chosen based on the likelihood the participants would contribute rich, qualitative feedback about what they thought and felt when evaluating teachers, and had experience confronting performance issues when necessary.

The findings of this study were as robust as the thoughtfulness, thoroughness, and honesty of the principals’ responses. From the standpoint of the researcher, principals’ opinions about values were as important to answering this study’s research questions as their factual evaluation decisions. Unfortunately there was not an agreed upon way to assess value judgments as they tend to be emotionally laced and specific to individual personalities and schools, whereas factual information could be evaluated by how closely it resembled the actual state of affairs. That being said, examples of disagreement amongst the participants were included in the findings to allow the reader to review dissenting opinions (Dalkey et al., 1972).
Chapter 4: Study Findings

Policymakers in Washington, DC and Denver have created new state accountability systems that significantly changed teacher evaluation. In part, these changes have affected how frequently teacher evaluations occur, how teacher effectiveness is measured, and how tenure is awarded or removed. This study focused on principals’ perceptions of those factors that influenced their evaluations of teachers’ professional practice (one half of a teacher’s evaluation) relative to how students performed on national, state, and local assessments (the other half of a teacher’s evaluation). Fifteen elementary, middle, and high school principals from one district were interviewed. The interview protocol followed a conversational approach reflecting the Theory of Planned Behavior (TPB) conceptual framework. The participating principals came from schools located in rural, suburban, and urban neighborhoods that varied in size, resources, and socio-economic status of the families served. All of the schools could be described as “traditional” public schools serving a mixture of neighborhood and school of choice students, with the exception of one alternative high school. The two major research questions were:

1) What factors do principals perceive as influencing and shaping how they evaluate teachers’ professional practice?

2) How do principals perceive the manner in which these factors influence and guide their performance ratings of teachers? In what ways did the principals feel these factors influenced their performance ratings of teachers?

Through their narratives the principals communicated what they felt influenced their evaluation tendencies and behavioral intentions, and even “vented” their frustration with a lack district level support while describing the challenges they’ve faced implementing the new evaluation system. Some participants were decisive in their responses while some answered
using pros and cons and were reluctant to take a position. The findings revealed the following: 1) principal-teacher relationships within the specific school culture was the main determinant influencing the extent to which principals engaged in critical evaluation of teachers, 2) the workload and job complexity of building principals influenced the level of detail and diligence they dedicated to their evaluation practice, 3) principals believed the overwhelming majority of PVSD teachers were effective or above and there was little they could do to improve or remove ineffective continuing contract teachers, 4) the new 5D+rubric was an improvement over the previous instrument and the professional development provided by the school district and University of Washington’s Center for Educational Leadership (CEL) enhanced principals’ efficacy in conducting teacher evaluations, and 5) district level leadership culture and support ultimately affected if, when, and how principals moved in the direction of creating improvement plans for teachers and/or pushing for transfer or dismissal.

**Perceived Factors Influencing Principals’ Evaluations of Teachers**

The factors identified across all respondents are clustered around key themes that emerged from analyzing the raw interview data and include personal, school, and district influences.

**Principal-Teacher Relationships**

**Individualizing evaluations.** Two personal factors that influenced principal decision making through the evaluation process emerged. The first of these concerned the personal background, ideas, and beliefs of the principals themselves, and the second related to the individual teachers they supervised. While the thinking and perspective of each principal was
unique, the factors they offered that influenced their decisions when evaluating teachers were consistent. There were no obvious differences in responses based on demographic factors such as gender, age, or ethnicity. There did, however, appear to be some variability in responses relative to the principals’ past experience confronting poor teacher performance that lead to a more cautious approach in dealing with cases. In general, if principals had not critically confronted the performance of a continuing contract teacher they were generally more positive and confident of their authority and ability to improve continuing contract teacher practice. An elementary principal expressed how little he could do about a teacher who was rated as ineffective this way,

They either can’t or won’t improve, and they’re often having a detrimental impact on their students. The worst teacher I ever had the great displeasure of working with I finally got out of my building after two years... I often think about her. I wonder if I really did any good because she didn’t leave teaching. They [human resources] moved her to another school, and they only moved her to another school because I told them I absolutely will not stop being honest with her and giving her feedback and pushing her to be better because she’s terrible. So they got to the point where they’re like, we’re tired of dealing with you so we’re just going to move her to another spot, and she taught for another five or six years before she retired. Thank God she’s retired now, but how many students did she negatively impact in those five or six years?

All the principals promoted a student-first ethic in asserting their responsibility to do the right thing and ensure there was an effective teacher in every classroom. Most principals went so far as to say their guiding criterion was whether or not they would place their own child in any of their teacher’s classrooms. As a middle school principal said,
The principal is ultimately responsible for every behavior of every human being under the roof that they serve... teachers, parents, and students. They can’t make every decision, but they’re responsible for everything. So for me it’s a moral imperative. Am I okay with my child in that class unsupervised by no one else but that teacher? If I can emphatically answer that with a thumbs up, I’m on the right track, but if I’m slightly doubtful then I have to do the right thing by everybody else’s child in that class. That’s what drives me. I’m not trying to get anybody, I’m certainly trying to help people improve, but for me evaluation is about -- have I allowed a group of students, 150, four periods a day, to be with a teacher that is not worthy of that trust?

Contemplating what’s most critical for them in the evaluation process as supervisors, the principals placed an emphasis on knowing who their teachers were as people. This was reflected in how they identified areas of strength and weakness when assessing teachers, and in how they went about fortifying the professional bond with their teachers. As a middle school principal said,

I look at all of them [teachers] through the individual lens, kind of the special ed. lens. In my own brain, every teacher is on an individualized education plan in some respects as much as I have capacity for… I talk with my administrative team a lot about who the runners are in the building and trying to keep the shackles off the runners and seeing if our administrative team can keep up with those people and shine the light on them, and if we can shine the light on the practices of the runners then there’s lots of other people who want to do those things. One of the challenges of evaluation is we have lots of good people that nobody knows they’re good.
Principals also considered the emotional health and stability of their faculty and their potential reaction to feedback. One principal commented, “People can learn skills, but you can’t necessarily change people’s personalities. That’s one of my biggest struggles.” An elementary principal thought about not only what feedback to provide, but also how his teachers would receive it and whether or not it would change the dynamic between them.

I have one teacher in particular who you even ask a question that’s just a genuine wondering and she’ll take it so much to heart that she’ll think she’s doing something wrong, and she’s such a good teacher that I find myself just focusing on positives with her, although that goes against my belief system as an evaluator. Then I have other people that kind of have the opposite attitude of bring it on I want to hear everything you’re thinking, if I don’t get something out of every time you come into my room you’re wasting my time. Those sort of people I feel like you can push more.

The principals also demonstrated a degree of discriminatory behavior on their part when they considered personal qualities and background in evaluation decisions (i.e., age, gender, personality traits like energy, enthusiasm, extroversion, and the ability to relate to children and get along with adults, experience, level of education, caliber teacher preparation program, licensing exam or applicant screening scores, etc.). For instance, in elementary schools where gendered power structures were more common, with men typically managing and women making up a larger percentage of the teaching workforce, qualified male teachers applied for teaching positions much less frequently than females and were subsequently harder to hire. So, a marginal male teacher might be retained and given more coaching relative to a marginal female teacher. Beyond efforts to hire and retain historically under-represented genders in certain grade levels and subject areas, principals also wanted to fill their buildings with teachers who delivered
effective classroom instruction and made a healthy contribution to the adult relationships in their
grade level teams, or departments. One elementary principal said,

The work ethic may not always be there. We had a guy who really fell short in that area.
He was just as lazy as could be. The team struggled with him. Kids are coming in at 8:45
and he’d be at your door saying “what are we doing today?” Didn’t really contribute,
people were frustrated with him, alienating his colleagues… and he was the nicest guy in
the world.

That teacher now works in a middle school. This principal was not alone in asserting his opinion
that a person’s character and work ethic isn’t likely to change much over time, and is not
something a principal can easily improve. There were also other, observable and unobservable
reasons, which principals deemed more or less persuasive in deciding whether to move from a
coaching relationship to an evaluative one, i.e. if the teacher had a history of success with the
prior administration or in another school, coached or had expressed an interest in coaching a
sports team or hard-to-fill extracurricular activity like a robotics club, or had political influence
in the district or community. While the merit of whether or not these reasons warranted a move
to an evaluation mode was debatable, what wasn’t was that principals thought about them.

Principals also said they prioritized working with new teachers when they entered their buildings
to familiarize them with the evaluation rubric, and as one elementary principal who approached
apprentice and experienced teachers via a differentiated process said,

I have teachers in my building who have taught for twenty-five years, and one [year]. So
I can’t approach everybody with the same expectations or mentality. What I found myself
doing as a newer principal was looking at my veterans going yeah they’re fine, I need to
worry about these people [new hires]. The flaw in doing that is that twenty year veterans
still want some growth. They still want to become better, they still deserve my full
attention in a post-conference, they still deserve a one hour formal observation where I’m
giving them written feedback. So that was something I changed very quickly in realizing
I can’t give all my attention to the younger ones, everybody needs it, they just need it in a
different way… everybody gets feedback.

That being said, there was a sense from the principals interviewed that some teachers simply had
what it takes, and others don’t, which reinforced the idea that high level teaching is some type of
mystical talent. As one middle school principal said, “My absolute highest performing teachers
are always willing to make change, but their locus of control is more internal, so I don’t have to
be as sophisticated in my maneuvering in how to get them motivated. They want the results.”

Principals, as middle managers, also recognized their own personal leadership styles
evolved over time. One middle school principal reflected on how his hiring decisions influenced
his evaluation decisions,

One of the biggest mistakes I’ve made was because of my ego. When I called the
previous principal for a reference check, he said “she was the biggest saboteur in my
school” out of over 100 staff members. I interpreted that as you didn’t do a very good job
getting along with that person, you didn’t create an environment that caused her to
flourish. I will. It was my ego; blind spot. The other one… I heard a rumor that this
person staged a walk-out at his previous school and was fired. I was also told by a parent
if you hire that teacher I’m not bringing my kid to this school, he’s amoral and unethical.
Throughout my career I’ve always believed you can’t ever give up on a kid, ever. When I
became a principal I carried that same belief system into the adult world, and I got really
good at convincing myself I could handle people who weren’t well. Then the true
behaviors started to reveal themselves. As long as I didn’t hold those two accountable everything was fine. As soon as I held them accountable, the poop hit the fan. In the final analysis it was my fault. It was a bad hiring job.

Throughout the study principals noted the influence of the hiring process with its ever-hopeful promise that they could find a more suitable applicant in the available candidate pool if a teaching position was vacated. This was especially true at the elementary level, where principals highlighted over 200 candidates apply for a typical position, which made it easier for them to maintain high expectations for their teacher’s performance.

**Probationary versus non-probationary.** Clearly, teachers’ employment status (probationary, or non-probationary) held significant sway over how principals approached evaluating them. Senate Bill 191 (SB 191) defined the word teacher as, “A person who holds an alternative, initial, or professional teacher license issued pursuant to the provisions of article 60.5 of this title and who is employed by a school district or a charter school in the state to instruct, direct, or supervise an education program (2010).” More specifically, the term probationary teacher in SB 191 refers to a teacher who has not yet completed three consecutive years of demonstrated effectiveness as determined through his or her performance evaluations and continuous employment, whereas a non-probationary teacher refers to a teacher who has completed three consecutive years of demonstrated effectiveness as determined through his or her performance evaluations and continuous employment. Over the course of the interviews the respondents viewed probationary teachers as essentially at-will employees with no job protections. This was most apparent when principals made hard line, no nonsense statements about their expectations and actions in deciding retention or nonrenewal. Principals used the words non-probationary teacher interchangeably with continuing contract teacher, or tenured
teacher. Whenever principals were talking about dismissing a teacher, they were typically referring to a probationary teacher. Conversely, whenever principals talked about the limitations of their authority, they inevitably referenced working with a non-probationary teacher. One middle school principal said, “I’ve had very powerful and positive conversations with probationary teachers, who want to make change. Their willingness to make change is much higher than a struggling teacher who is continuing contracted.” All the principals identified energy, enthusiasm, passion, and a can-do attitude as the attributes they wanted to see in their faculty members, and they more often found these characteristics and behaviors in probationary teachers. They felt soft skills were harder to teach and more important than hard skills, and the ability to accept feedback was more important than experience. They believed younger, newer teachers to be easier to work with and coach, which made them a better overall fit for their buildings. Thinking about how a teacher’s employment status influenced his summative evaluation decisions a middle school principal said,

I have not been afraid throughout the years to non-renew a probationary teacher; that has not been an issue for me at all. But I have been hesitant to rate a continuing contracted teacher less than what I think, and I have improved in that area. That’s something I have not done a very good job of early in my career.

Principals understood not all teachers required the same amount of supervision or support. As a result, they differentiated their application of the 5D+ evaluation rubric to teachers contingent upon their standing as probationary versus non-probationary. One middle school principal commented,

I can give them [effective non-probationary teachers] some overall good information and they’re appreciative of it. Different experience with a brand new teacher, or a struggling
experienced teacher. The brand new teacher is hungry and more vulnerable because they are an at-will employee that can be non-renewed. With those folks I use the rubric and this evaluation process to a T. I give them as much feedback as I can, and they don’t hold it against me as a trust issue. They want to do well, please me, and hit the mark. They’re seeking continuous improvement. It’s the teacher who struggles with any feedback less than effective throws them into a tizzy. It becomes all of a sudden a trust issue with the administrator, and for years I shied away from it because intuitively I knew it would hurt the relationship, especially since they’d came from another school they’d been at for twenty years and all their evaluations were great.

Principals shared their understanding of the additional job protections that historically came with a teacher beginning his or her fourth consecutive year of service and moving into non-probationary status. They knew that once a teacher became non-probationary, they were theirs forever. A high school principal declared,

It’s worthless to even discuss someone who is not probationary even though HR will say, “No, you can put him on a performance plan.” When was the last time you saw that work? I’ve got two or three individuals that I would love to do that with that I’ve met with HR ten, eleven times about and I don’t have any confidence anything will change. The only time I’ve ever seen people lose their jobs is they do something illegal, do something with a kid.

However, the rules changed under SB 191 and tenure protections are now not awarded to a teacher merely beginning a fourth year of service, but only to those receiving three consecutive effective or above summative evaluation ratings. So, a principal could rate a teacher partially effective any year within that teacher’s three year probationary time period and it would start that
teacher’s three-year probationary clock all over again. Although very few principals interviewed thought this change would impact the number of probationary teachers they moved to non-probationary status, there was a sense from principals it would lessen the stress historically placed on the importance of the third year (of probationary status), and give principals more freedom to make long term, less permanent staffing decisions. A middle school principal commented,

I’ve seen three times in my time here where a teacher’s behavior changed in the fourth year, after performing very adequately the first three years and getting continuing contract status. I don’t know if it was caused by their continuing contract status, that they played the game until...

School Level Factors

The principals interviewed worked in schools with enrollments ranging from 150 to 1,965 students. School size did not appear to influence principals’ evaluation decisions. However, other site specific factors such as workload, climate and culture, and parent perceptions and opinions did.

Workload. Workload demands emanating from district and state directives, as well as, time limitations resulting from daily interruptions at the sites dominated principals’ interview responses about how they perceived the manner in which factors affect their teacher performance ratings. Overall, principals believed state policy makers were out of touch with the job complexity and workload of today’s building principals, expecting them to successfully evaluate every teacher every year on thirty-seven indicators and run a building. One high school principal confirmed, “There are teachers new to my school that I’m positive I haven’t been in their class
yet, and a couple I didn’t even hire. I’ve gotten so far as knowing their name and their discipline, but other than that I don’t know.” Another high school principal noted, “We’ve increased our workload by 200% with this process and yet there is no additional staffing, at least not for the secondary folks. We’re struggling getting in to see teachers even just for walk-throughs.”

Principals attributed the workload increase to a number of factors. First, all licensed teachers (probationary and non-probationary) are now evaluated every year. In the past only probationary teachers received annual evaluations, while non-probationary teachers were evaluated on a three-year cycle. These non-probationary teachers could account for two-thirds or more of a typical faculty, hence the principal’s reference to a 200% increase in workload. Second, the rubric itself is more comprehensive than the previous document with thirty-one instructional indicators to assess. Third, the iPad and cloud-based technology used to collect the evaluation data required new learning and processes for principals. Fourth, elementary principals expressed a stronger desire to visit classrooms the recommended 4-6 or more times per year to complete formal observations (30 minutes or more with a written record and post observation conference), informal observations (15 minutes or less), and informal walk-throughs (2-5 minutes). One elementary principal asked his teachers when they were feeling overwhelmed by the thirty-one instructional indicators on the rubric (not counting the six associated with the dimension of Professional Collaboration and Communication), “What would you take off?” and they couldn’t pick one. They believed all the indicators had merit.

However, secondary principals faced a different set of circumstances. Those principals noted that what made evaluation difficult is simply finding the time to get into classrooms to complete the two required formal observations and give teachers feedback afterward due to the
employment demands and amount of distractions they encounter in a given school day. As a middle school principal noted,

We do a pre-conference for 15-30 minutes to go over lesson plan design to code things on the lesson plan. Then we do a 74-minute visit typing the entire time, then probably another hour, hour and a half to get it ready to talk to a teacher, then we talk to the teacher for 45 minutes to an hour, then I have to go back and change stuff because there’s no spell check, and all that bullshit, I’m sorry but that’s a big fat waste of time, so we send it out and if they have questions we come back. So it could be anywhere from 4 hours to 5 [hours] per person.

Both the PVSD human resources department and the Pioneer Valley Education Association (PVEA) have recognized this issue, and in response, approved allowing principals and continuing contract teachers at the sites to waive the required second formal observation for the 2014-15 school year by mutual agreement. A middle school principal continued,

For a highly effective principal to have to do seven learning walks a day, plus formally evaluate everyone, it’s impossible. I did it last week, and I got nothing done in terms of kids or people... I keep thinking, wow, is this how it’s going to end [reflecting on her retirement next year]? I’m not going to be with people anymore, I’m going to be with technology and rating systems… When kids are asking, “Can you please come without your iPad?” that really bothers me.

Principals blamed the obligatory documentation and new technology for pushing them into a management focus. A veteran principal declared, “The work is overwhelming, literally. I use that term purposefully because the amount of paperwork, study, typing, record keeping, goal setting, and the SLO and SOO process...”
The different perspectives presented by the elementary and secondary principals were notable, and indicated just how dissimilar the job is at the two levels. The elementary principals noted an exclusive focus on instructional leadership and used the new evaluation process to support that priority. However, the secondary schools’ larger student enrollments, numerous parent groups, broad-based athletic programs, extensive extracurricular offerings, variety of musical performances, and daily student management and discipline demands dictated the amount of time their principals could engage in evaluating and improving classroom instruction. Clearly, secondary principals did not see a high degree of value in spending the vast majority of their time in classrooms with teachers who they have been evaluating for many years and who they believed are effective or highly effective. They were more likely to use this process to evaluate struggling continuing contract teachers and probationary teachers, and provide one-on-one coaching. While all the principals felt like they’re busier and working more this year than last year, one elementary principal contradicted her colleagues who complained about the workload issue,

That’s what our work should be. I do not look at it as workload. I look it as what I love to do, what I need to be doing. I shouldn’t have to be restraining special ed. kids because I don’t have enough special ed. staff. My job is coaching teachers and improving teaching and learning in my building.

The principals tried to mitigate the workload mainly through delegation to support staff, if they had any. For example, one elementary principal’s office manager scheduled all of her formal observations with teachers ahead of time and as a Title One school had an in-building instructional coach to help coach teachers on elements of effective instruction. Another elementary principal summed up how the new law has taken a toll on PVSD principals this way,
As much as I like in theory the idea of the new evaluation process I will say this has been my most difficult year, my tenth year, at the same school in the same role. You’d think it wouldn’t be, probably first year, maybe second year as you take on issues, maybe even third, fourth year. I feel like the process is taking a lot more time.

**Climate and culture.** Principals cherished the importance of a healthy school climate and culture and identified the faculty’s reaction to their evaluation judgments as a substantial influence on their overall decision making. Principals were keenly aware that possessing a strong culture must precede changing structure. For instance, principals reported teacher anxiety was elevated from something as simple as the evaluation instrument ratings changing from three levels to four with the adoption of the 5D+ evaluation rubric. An elementary principal shared some of her teachers said, “‘I’ve always gotten the highest rating,’ well the highest rating was from satisfactory to rock star, one big band that was a big shift.” Principals from all three levels expressed angst knowing they are required to keep personnel matters confidential, yet they cannot control what and how teachers informally report out evaluation information to their peers. They simply have to hope their reputation and the building culture is strong enough to withstand any negative allegations. The different expectations of principals and their teachers around communication made maintaining relational trust amongst the staff complicated. One middle school principal asserted,

I’m unique in the sense that I’ve hired everyone in this building… so I’ve got it the best it’s going to possibly be. I didn’t inherit anybody, so it can’t get any better. I think what people would say is my expectations are very, very high, and it’s not mine only it’s our expectations to work here… We have great people and they know when someone is not cutting it. So, I always think of staff and say you know if you are upset with me that’s
fine, and when have I ever done anything that was not in the best interest of the future of this school? Name a single time. Sometimes this stuff has to happen and I’m the person to do it. Most often people would come in and say I knew it, I knew it.

Principals were conscious of the tension the new evaluation system created amongst their teachers, so some prioritized enhancing climate and culture by getting their teachers to see the new system from the balcony versus the basement in accordance with a growth mindset (Dweck, 2006). A high school principal said,

The increased accountability has taken its toll on our staff members as far as stress level… ours are always kind of on edge. It’s difficult to sit down and have a conversation if we’re seeing things in a classroom or hearing things from students or parents… they can’t relax and say, “Yeah, okay, I can see that and how to get it fixed. . . ” the increased stress has been noticeable.

In the final analysis, principals viewed their building’s climate and culture as a product of the collective attitudes of the teachers who worked in them. Principals reported frequently assessing the potential negative impact of a decision on their school’s culture. They saw how their annual evaluation decisions and previous hiring practices impacted the way their buildings felt to students, visitors, and staff. Referencing the negative effect of a teacher who transferred to his building one middle school principal recalled,

A teacher came to me after being in the district for fifteen or twenty years and had always gotten stellar evaluations, and then all of sudden I’m giving her feedback about her practice and as soon as I told that 25-year veteran teacher she needed to cooperate better with her peers and I would give her support, she refused to accept that. She did not see it as a performance issue, she saw it as a conflict issue. As I traced her career back through
the years it [the conflict between her and her colleagues] was a pattern in every school she had been at. When it would get too hot in those schools she would move. That was at least the theory of the two or three principals I talked to.

**Parent perceptions and feedback.** The geographic location of a school, or more specifically, whether or not that school served a more affluent neighborhood demographic, factored into principals’ perceptions. As one elementary principal remarked,

I’m in an environment that’s extremely high pressure, high anxiety from parents, and it’s too much for some teachers. I have had some teachers leave willingly, just saying love you guys, love it here, I need to get into a different kind of school, and they thrive. They appreciate being with kids in poverty and not having parents beating down the door everyday asking them how they’re teaching math and why they sent that homework home. That’s just the kind of stuff we deal with out here, and it’s not for everybody.

Most of the principals were cognizant of their parents’ desire to share their opinions when they did or did not like a teacher either directly through a conversation, or indirectly through an anonymous survey. Principals also shared their assumptions, and biases, about schools in other parts of town. An elementary principal stated,

My parents have extremely high expectations and have no problem telling me their opinions; good, bad, or indifferent. I get high turnout on teacher surveys. I get very honest feedback on my principal survey. Which is part of the reason I feel it’s my duty and obligation to make sure I’m upholding high expectations here. I don’t want to hear it from the parents, but I also have the parents backing when I nonrenewed a teacher, not my only one, one of many. I had several parents contact me saying, “Good move, glad to see that person won’t be back next year.” Heard that in the crosswalk one day… a dad
walked across the crosswalk very nonchalantly and said, “Glad to see Mrs. So And So won’t be back next year,” just kept on walking, “have a nice weekend Ralph.” Overtly you might not see it. “Oh my gosh I’m so sorry she’s leaving, we love her, she was awesome,” but internally those are the people coming to me saying “she needs to go.” Principals acknowledged the annual multi-source anonymous survey results from parents and students were helpful in delivering feedback to teachers, as it allowed them to hold a mirror up to teachers and say, “I’m not the only one who sees this, your parents do too.” However, that formal process did not occur during 2013-14 as the new district-adopted student management system, ironically called Synergy, was not compatible with the old multi-source online survey technology. Nevertheless, most principals still gathered 360 degree feedback at their sites in the absence of that district-generated multi-source survey. An elementary principal remarked, “I survey my staff internally about their colleagues, and we do student surveys as well, those are all really helpful.” Principals perceived that, in general, parent input did not dictate how they evaluated teachers; it simply validated what they already knew. Thinking about the level of parent support at her middle school, which did not serve a large portion of affluent families, principal contended,

No parent here questions what we do. I don’t think I would do well with that entitlement piece… someone calling a lawyer when a kid is sent out of class or anything like that. I don’t even understand that at all, and don’t really want to.

District Level Factors

District factors were perceived by principals as influencing their evaluation decisions. These included the professional development support they needed to implement the 5D+
teaching and learning evaluation rubric, human resources (HR) support to nurture teachers’ instructional capacity and confront poor performance, instructional technology support when it didn’t function properly, and research and evaluation support to generate accurate growth ratings for teachers.

5D+ evaluation rubric. Principals noted that their performance ratings of teachers were influenced by the adoption of the 5D+ evaluation rubric, as it changed the way they approached coaching and evaluating teachers. There was consensus among the principals that the new 5D+ teacher evaluation rubric described behaviors better and was a vast improvement over the former instrument. Principals liked that it was predicated on a coaching model and lessened the inter-evaluator disparity in buildings with multiple administrators. As one middle school principal said, “I prefer that teachers are rated as highly effective and effective. I like that language, rather than satisfactory and needs improvement.” Principals believed that the rubric could improve the collective pedagogical capacity of the faculty by providing principals with a common language to give more actionable feedback to address issues with teachers they may have avoided in the past. The rubric essentially added teeth to an evaluation system that didn’t have any before. That middle school principal continued, “I had grown to develop a bias on who was effective and who was not. What the rubric caused me to do was examine a little bit more carefully and honestly, through words and descriptors, am I seeing these things?”

Some principals shared they have had increased difficulty rating teachers effective, or highly effective in an indicator if they did not see evidence of that indicator. The sheer number of dimensions, sub-dimensions, and indicators in the rubric also made the evaluation process difficult. As another middle school principal advised, “I believe focusing on six dimensions is too much. I’d rather focus on one dimension and really nail it.” Principals provided these
comments knowing if they interpret the rubric literally, provide reasonable observable feedback, follow the inquiry process and complete the required observations within the timeframes outlined in the Employee Agreement, the content of their evaluations is not grieveable by the teacher or teacher’s union.

**Professional development support.** Principals said the significant inquiry-based professional development they received in the year prior to (2012-13) and first two years (2013-14 and 2014-15) of the new evaluation system roll out enhanced their sense self-efficacy. Principals liked learning about only one or two dimensions out of the six initially, e.g. purpose and student engagement. This made for a more focused and targeted study of the rubric language. In the process it became clear to principals that the level of sophistication and leaner centeredness increased from the left to the right side of the rubric as an evaluator stepped up the band descriptors from ineffective to highly effective. Principals also said the rubric enhanced their teachers trust in them, especially when they were evaluating teachers in content areas they admittedly were not familiar with, like music. A high school principal remarked, “The rubric really does create a science of teaching.” Principals mentioned each school was directed to evaluate teachers only on three of the six dimensions in the first year implementing the new system (2013-14) – purpose, professional collaboration and communication, and a third dimension chosen by each school site, but in year two teachers were assessed in all six dimensions. Principals were also told to only formatively *code* the indicators they observed in formal observations, and not formatively *rate* them, meaning, check each of the thirty-seven indicators they saw evidence of in a particular lesson, but do not assign a 1-4 judgment. Since principals were expected to rate each indicator for each teacher in three of the six dimensions
summatively at the end of the year, this created inconsistency across principals. Reflecting on his progress administering the 5D+ rubric in the second year, a middle school principal said,

I’ve got a teacher that I consider an effective teacher, and I was in her classroom doing a formal observation the other day, and I went through all 37 indicators and she hit the mark on every single one. In her post-conference she said when she was doing her administrative internship last year she got deep into the 5 dimensions to the point where she has them memorized, like the difference between this and that, and even though I would rate her as effective it felt like she did a little dog and pony show for me and hit the checklist. She’s good.

All of the principals agreed the most helpful professional development they received was the three-day inter-rater reliability training that began the 2014-15 school year. Administrators watched videos of classroom instruction, rated the teachers, and compared their observations. Over the three day training a baseline score was required by PVSD to “certify” each administrator had met the minimum standard to evaluate teachers using the rubric. This condition of a passing score offended some principals, as one middle school principal declared, “We’re licensed by the state to evaluate teachers and I don’t need the pressure of if you don’t make it you can’t evaluate teachers because if I pushed that we would win.” Even so, principals reported the professional conversations they had during this training about such concepts as “discipline-specific habits of thinking” and interpreting the rubric’s frequency words, such as “occasionally,” “frequently,” and “consistently” were thought provoking. On the whole, district principals felt the professional development they received helped them be more consistent, regardless of the grade level or content of the teachers they observed, and were enthusiastic about the district’s decision to align all departments around the 5D+ terminology and framework.
That same middle school principal conceded, “The inter-rater reliability training I complained about because it was set up incorrectly actually was good.”

**Human resources (HR) and cabinet support.** Some principals felt inadequate prescribing specific strategies to enhance instruction, whether directed at an effective twenty-five year veteran or a partially effective ten-year veteran. Regarding low performers, principals were unsure about exactly what they should be doing, when they should be doing it, and/or how they should be doing it. The human resources department advocates “progressive discipline” before creating an improvement plan, and moving in the direction of dismissal, but no principal said they ever have had a representative from HR walk them through what a progressive discipline process should look like. So, principals were conditioned to confront poor performance on their own and not expect assistance from HR. An elementary principal described her approach to progressive discipline as,

> Having the hard conversation, pressing for compliance if you need to, or following through on expectations. I’ve changed grade levels for people just to make ‘em mad, and I’ve had people leave because of that. I’ve gotten somebody to retire, which I know sounds awful, but this wasn’t the place for em’… they were three years from retirement. I can’t say I’d press real hard to get rid of em’, just make life hard for em’, and maybe get them into another building where it would be a better fit. Because sometimes people really do perform differently in a different environment.

The principals acknowledged having anywhere from three to twelve teachers they would label as performance concerns in their respective buildings, depending on the size of their faculties, which ranged from 16 to 120. They also confirmed the dimension of concern was almost always Professional Collaboration and Communication (PCC), or PCC combined with another
dimension. When principals spoke about their experience confronting mediocre performance they recalled the teacher being supported by their building union representative, district union president, area uniserve director, and sometimes even that teacher’s personal attorney. In contrast, the principals said they were assisted by a human resource specialist who introduced him or herself as “just an observer” to monitor the process. In situations where the issue rose to a higher level and the Executive Director of Human Resources joined the principal, the principals felt an increased sense of support. Numerous principals communicated PVSD is currently on its sixth executive director of human resources in the past ten years. An elementary principal whispered,

The one person I felt most supported by ever at central office is gone [former HR director]. He was fired, and I don’t know all the details of it, but I’d be hard pressed he wasn’t gotten rid of because he was targeted by the union. Don’t know that for sure. Having some of the information, and knowing some of the background, it’s extraordinarily hard to come to a different conclusion. That makes you pause and think, should I really be pushing as hard as I want to, I do have a high performing school; our kids are getting a good education? I don’t know.

A middle school principal stated he has a teacher on paid administrative leave for insubordination and the school board will be voting on that teacher’s dismissal at an upcoming board meeting. This principal worked with the last three HR directors over a four-year period regarding this individual, and had to orient each new HR director to the situation and manage their sometimes contradictory approaches to improving the situation. Regarding giving feedback to a teacher teetering between partially effective and effective, with continuing contract status and the support of the union, a principal said,
I rated a teacher who has been teaching over twenty-five years partially effective in one sub-dimension out of thirty-seven and her message now to the rest of the staff is I am not to be trusted. I like to get along, and I think trust is foundational to being collegial and being a team. It cuts to the core of what is at the foundation of a healthy organization. So I’ve had to do my own cognitive work around is it really a trust issue? No it’s not. When a person who is given feedback in only one sub-dimension out of thirty-seven that’s below their expectation, they’re not open to continuous improvement, they’re even willing to go so far as polarize the staff into camps around trust. So, that’s where this can lead, and I’ve been willing to go there with a few people this year, and I don’t feel bad about it at all, I wish I started it sooner.

Principals voiced their respect for the current superintendent and her desire to prioritize her relationship with the school board. However, they also acknowledged their frustration with some of the superintendent’s cabinet members. As this principal remarked,

The problem with that position [the superintendency] is she has to have people under her that are strong, that are leaders, that are fearless, that are courageous, that are kind, that are role models, you fill in the blank and that ain’t happening… she has the wrong people there because I don’t see myself in any of those people, nor do I feel support from them, ever. Mainly we have an avoid screwing up attitude… I’ve got to soar and I’m walking with my feet in the mud… Don’t do this, don’t do that, don’t stand up, don’t go to the bathroom, don’t color outside of the lines, don’t be innovative, don’t lead… manage, manage, manage.

The opinions of the principals’ direct supervisors were perceived to be a small factor in principals’ decision making on teacher evaluations. A few principals remembered being asked by
an assistant superintendent at a principals’ meeting three or four years ago to write down the names of the teachers in their buildings with performance concerns on an index card and turn it in. They said little came of it other than adding more pressure to an administrative culture of high accountability. A high school principal shared,

Our current asst. superintendent is a ball buster and expects principals to weed people out pretty quickly if they’re not doing the job. At least that’s the impression I’ve gotten. I’ve felt that pressure. In our building we’re trying not to take that approach, but we have that piece hanging over our head… either coach folks out or put so much pressure on them they resign and quit. I don’t sense that from our current superintendent and so that’s a way to kind of balance things… when we were told who are your weakest teachers, come up with a list, hand it in. In my mind that’s your hit list. I gave her my list of about eight people I had struggles with. I’ve got that list in my desk still. Those are the ones I spend the most time helping improve. I’m in their classroom more; I talk to them more, because I want to see them be successful. Are there a couple I’m still working on as far as they’re not right for the profession, absolutely, and we’re getting close on a few that will retire in the next year or two so how much time do you spend on that?

Principals were also cognizant of how well they wrote up summative teacher evaluations, as their boss’ signature was required on every licensed summative evaluation as the next level manager. Three of the elementary principals noted their previous assistant superintendent favored summative evaluations that were heavy on the narrative component. One secondary principal said he received a call from his current assistant superintendent a couple of years ago directing him to include more content. In general, principals perceived that the general expectation was that summative evaluations should be written professionally and be legally defensible. As an
elementary principal said, “Regardless of who my boss is if I have a teacher here who’s doing things that are harmful for kids I’m going to try everything in my power to get rid of them and if I can’t I’ll probably quit. I’d like to think we’d all do that, but I don’t believe that to be the case.”

A middle school principal said,

I’m not too mindful of anybody out of the building and probably to a fault… I feel like I’m doing what’s effective for my people. The energy in my brain goes to the individual I’m engaged with at the time… knowing who I’m talking with and knowing how hard I can throw the pass.

A few principals responded that one of the assistant superintendents was out of touch with the intricacies of teacher evaluation procedures at the site level. Some cited other district personnel whose opinions influenced evaluation outcomes, like the induction program coordinator who stressed showing up to meetings on time, holding people accountable, and making sure teachers know their place over building capacity and promoting growth. A principal said,

I don’t think that there’s a whole lot of people overseeing the process unless there’s shit hitting the fan, and I don’t necessarily value people’s opinion that aren’t in my school on a daily basis. I had a P3 who I sent in the evaluation and I got contacted by the head of human resources who said sorry I can’t agree with you on this. This person is going to go into their fourth year, but we’re going to keep them probationary, and it was because this person got on the wrong side of the person who runs the induction program and I supported my teacher. He didn’t dot all the Is and get his three ring binder in the correct order and he took it upon himself to use the new teaching standards instead of the old standards for his induction and he got chastised.
Principals said they were mindful of the teachers’ union. They noted that although they thought its leadership has not been as participatory as it could be, they welcomed it into their buildings. An elementary principal remarked, “This is a heavy, heavy PVEA building and I don’t have any problem stepping up and doing that part of the job because of the twenty-five faces multiplied by how many years by how many scenarios.” Another elementary principal confided,

I don’t trust the union at all because I know that the union president currently in our district was a highly ineffective teacher. He would have been number two that I was coaching out of the profession. If he had stayed in my building he was next, and now he’s leading my teachers union, which kills me… It’s more about protecting teachers versus looking at practice... The union in my case protected a very ineffective teacher, went to bat for him, not okay.

**Instructional technology support.** Referencing technology support and the challenges of completing all walk-throughs and observations using an iPad and the new cloud-based GoObserve application a principal said,

I’m not a very prolific keyboardist. I find it difficult to have an effective evaluation because I’m capturing what I hear for the most part not what I see because I’m staring at the screen and I find that disrespectful to teachers... especially with the 5D rubric’s foundational beliefs that teaching is a highly complex endeavor… and I’m capturing half of it and that’s the auditory half, not the visual half.

A middle school principal echoed this frustration, “I was in the middle of an evaluation today and it went off. It was the third time. Going back to the start is cumbersome, and the looking down, I like to see and celebrate and not miss something.” Principals also experienced malfunctions including iPad cameras not working, observations not saving correctly to the cloud,
or observations getting stuck in the iPad application only to be sent months later to teachers. An elementary principal offered,

I think that’s completely ridiculous and I hate it. The scripting process and the coding process… I would be willing to do the workload if I thought it was valuable. The first time I printed out a report from GoObserve it was eighteen pages long and my teacher brought it to me and said this is not you, and she like wiggled it.

**Research and evaluation support.** The respondents also questioned the reliability of the growth scores calculated by the district for their teachers. Principals admittedly couldn’t explain the Hierarchical Linear Modeling (HLM) that computed the scores, and ultimately had to trust the system, which used at least two years of student performance evidence to determine a teacher’s growth rating. Principals speculated whether the spirit of SB 191 was to sort or support teachers. They were under the impression that all of the schools likely had numerous highly-rated teachers with growth scores of three or four if they had the requisite two years of data available. For the few teachers who did not earn high growth scores principals speculated how teachers might game the system to improve their ratings. An elementary principal surmised, “If I was a teacher again and thinking of nothing but my growth score, what I would want to be is a fourth grade teacher at the school with the worst third grade teachers, because you’d get a huge growth score.” Principals also took into consideration the difficulty of generating a growth score for teachers in unique departments or programs, such as, integrated services (special education), teen parenting, or online credit recovery, but didn’t have adequate stratum one or stratum two data. These teachers, like new-to-PVSD teachers who did not have two years’ worth of standardized assessment data to “flag out” in stratum one or two, had to create a Student Learning Objective (SLO) to generate a growth score, which some saw as a massive waste of
time. The flag out process refers to PVSD’s three-tiered value-added model in which teachers can be “flagged out” at either stratum one or stratum two based on their students’ standardized test performance before having to go through the SLO process (stratum three). One elementary principal said this about discrepancies that emerged over ratings derived from standardized test data (stratum one and two) and the SLO process (stratum three),

> Everybody in my building rated as highly effective based on their student growth [from the SLO process], scored their own results. Nobody in my school that has state standardized measures for TCAP [from stratum one and two] was a four. Now we’re going to win the John Irwin Award again this year, we’re going to win the Governor’s Award again this year, and I just got back from DC yesterday after winning the Blue Ribbon Award. Not one of my fourth or fifth grade teachers is rated highly effective, but a first grade teacher who can score herself on the DRA2 can be highly effective pretty easily. I’m not suggesting anybody is saying I’m going to grade these inaccurately so I score highly effective, but I know I was guilty sometimes of thinking oh I know what they meant. I’m going to give them a three instead of a two and all of a sudden you have a kid pushed slightly into a category where they probably wouldn’t be if it was an outside entity doing the scoring.

This principal was not alone in pondering whether or not teachers who scored their own student results through the SLO process were more likely to get a three or a four than those who did not. Building principals also contemplated how teachers’ growth ratings correlated to their professional practice ratings. As one observed,

> I struggle with the number, and as much as I do think it is statistically reliable and valid. When I see my eighth grade history teacher getting the highest rating [4] because it
comes from statewide writing growth scores and he doesn’t teach writing, and I give him a 2 or a 1, it’s not very defensible.

A high school principal expressed his aggravation that the flag out process was supposed to reduce principal workload by having only about 10-15% of teachers completing SLO’s. However, in his building that number was closer to 50%. He said, “We have teachers saying, ‘What do you mean I didn’t get flagged out?’ You know you’ve got some rock stars that have SLOs and you know you’ve got some teachers that need some work and they did get flagged out.” While he shared the reasons a teacher might not get flagged out with the teachers who complained, i.e. there were not at least four teachers teaching that course in the district, their N count was too small, or their growth was not consistent enough, the teachers still questioned the legitimacy of the measurement. However, an elementary principal presented this contradictory feedback,

When we look at the fifty percent measure based on student growth there’s a real high level of certainty in my mind and in our statistician’s mind about the reliability and accuracy of the data we get from fifth grade for sure, from fourth grade for sure, and from second and third grade. I think the problem I’ve seen with our district and probably even more pronounced in other districts are those people who don’t fall into the obvious categories of standardized test data that you have for them, like my kindergarten or first grade teachers, or my specials teachers.

Accordingly, principals had divergent perceptions of which measure was a better gauge and questioned their own efficacy in evaluating performance. An elementary principal said,

I think student performance growth should absolutely be a part of a teacher’s evaluation and a part of my evaluation, and most people would disagree with that. If your main goal
is to help kids learn and you have measures that have proven to be accurate to show how much growth they’re making towards certain targets why wouldn’t you use that? It’s a whole lot more fair measure than me going into a classroom and arbitrarily saying hey you’re highly effective or you’re not, you’re partially effective or ineffective. In the past the evaluation system has been pretty much my opinion… now if you happen to get the right principal, I think that model works just fine.

**Isolation and lack of autonomy.** As principals talked about the perceived influences on their evaluations of teachers, a recurring theme was the degree of isolation they felt in doing the work. All principals appreciated having at least one part-time or full-time assistant principal (AP) to discuss evaluation practice, especially in the elementary schools which began receiving staffing for APs in 2014-15. Principals knew the evaluation decisions in their schools rested squarely on their shoulders and they were forbidden to talk publicly or privately about personnel issues. An elementary principal noted, “I know I’m making the right decisions for kids, and that’s what it comes down to.” Yet principals’ understood their evaluation decisions were not interpreted in isolation as teachers freely talked with the union, their peers, parents, and sometimes even students. As a result, principals, especially at the high school level, found themselves thinking about how to field questions from parents and students about what happened to a popular teacher might affect their decisions.

In addition to isolation, principals noted a lack of control or autonomy as something they believed influenced what course they should have taken, if any, to improve teacher performance. Principals professed a gap between the amount of authority they thought they needed and the amount they actually had for such tasks as designing and administering improvement plans, and
discharging or transferring inept teachers. An elementary principal expressed the perceived limitations of his control and how frequently principals accepted mediocrity by saying,

They’re not doing a God-awful job, they’re doing okay, and they’re tenured. I know what that means if I take them on. I’ve had my job threatened more than once by some pretty powerful people and it’s because you take on teachers who happen to be not performing at the level they should be… Yeah I’m going to be looking for a job here soon myself, not because I did anything wrong in the evaluation process, unethical or illegal, but because I told them the truth. I told them honestly where they were currently functioning and they didn’t want to hear it, and they have pretty powerful allies with the teachers’ union.

The teacher this principal was referring to was ultimately transferred and the receiving principal was told to manage that teacher until she retired. Principals perceived these barriers as the product of district policies and procedures, the negotiated agreement, and SB 191, which clearly stipulated principals do not have the authority to hire or fire teachers. They can only recommend hiring, renewal, nonrenewal, and dismissal. Ultimately, these decisions resided with the board of education.

Another example of principals’ shrinking autonomy was the possibility the district would impose averaging, or weighting each of the 6 dimensions and 37 indicators to calculate teachers’ summative evaluation scores. There are as few as 5 sub-dimensions in the Purpose dimension and as many as 7 sub-dimensions in the Curriculum and Pedagogy and Classroom Environment and Culture dimensions, and averaging was viewed as an unnecessary restriction on principals. Currently, in year two of the rubric, the 37 individual indicator and 6 overall dimension ratings are at the discretion of the principal. Many felt averaging would reduce the impact of principals’
professional judgment on summative evaluation decisions by deferring to a calculation thereby making those judgments appear more objective, like teachers have done for decades when a student received a 79% versus an 80% in a class. However, some principals perceived this philosophy of deferring to a higher authority, or a district directive to average the six dimension ratings, would diminish principal voice and the expectation that principals know the rubric, know their teachers, and use the descriptors embedded in the rubric to correctly place teacher performance on the continuum. Concern was also expressed that averaging the indicator ratings would drive principals to “game the system,” and assign a rating in an indicator that is not authentic to ensure the overall dimension rating becomes what he or she wants it to be, or taken a step further, assign a rating in a dimension that is not authentic to ensure the summative rating across the six dimensions is what he or she wants it to be. Principals contemplated whether an ineffective overall score in one or two dimensions should trump effective scores in the other five or four dimensions when it came to the final professional practice rating? What if a teacher ended up with three partially effective ratings and three effective ratings? Should he or she be averaged up to effective? Principals wanted to retain the authority to assign a summative evaluation rating of partially effective or ineffective, using the rubric and its marked levels of performance, if the teacher’s lack of professional collaboration and communication warranted it, even though that teacher may have been effective in other dimensions. Overall, principals did not want an average summative calculation using the mean, especially those principals who had adopted standards-based grading practices in their buildings, as they knew that the least effective measure of central tendency (mean, median, and mode) was the mean.
Chapter 5: Summary, Conclusions, and Recommendations

Colorado Senate Bill 10-191 (SB 191) passed in 2010 in support of the U.S. Department of Education’s current Race to the Top program and earlier No Child Left Behind Act, mandated increased teacher and principal accountability by affirming improved teacher evaluation will advance K-12 schooling. After adopting a new evaluation instrument called the 5 Dimensions+ Teacher Evaluation Rubric (5D+) and creating a new value-added model (VAM), Pioneer Valley School District (PVSD) principals began annually evaluating all licensed staff in their buildings in 2014-15 and advanced their professional learning to use these new tools and meet the law’s requirements. This case study explored the dynamic nature of building principals perceptions, interpretations, and actions in the context of these changes in teacher evaluation. The central research questions were:

1) What factors do principals perceive as influencing and shaping how they evaluate teachers’ professional practice?

2) How do principals perceive the manner in which these factors influence and guide their performance ratings of teachers? In what ways did the principals feel these factors influenced their performance ratings of teachers?

In the face of increased accountability on principals and teachers, new evaluation documents and technology, and the significant professional development that accompanied this law, the principals interviewed in this study still largely approached the evaluation process as they had in the past. In other words, despite the government’s best efforts, the percentage of teachers annually rated highly effective, effective, partially effective, or ineffective by principals for the most part remained the same for year one and appeared to be holding true in year two of implementation. Research (Hanushek, 2010; Weisberg, et al., 2009) has shown that less than one
percent of teachers nationwide are rated ineffective annually, and this has been consistent with PVSD results.

Overall, the principals’ narratives illustrated their thinking behind the evaluation process and how and why they chose to address, or ignore teacher performance issues. Their perceptions depicted a reality in which probationary teachers were evaluated more thoroughly and critically than their non-probationary counterparts. More specifically, principals felt the rubric training was helpful, their own expectations and the school culture primarily influenced how they rated teachers, the overwhelming majority of teachers were effective or above and ineffective tenured teachers were difficult to improve or remove, district level leadership support affected if, when, and how they responded to performance concerns, and evaluating every teacher every year added significantly to their already challenging workload.

Although the district embraced SB 191 as mandated and summative evaluations changed as a result of the new tools, the values, beliefs, and goals principals brought to the day-to-day evaluation process remained the same. As a middle school principal recognized, “Top down when necessary, bottom up when possible… this came from the top down, and that’s not as easily embraced because it feels like it’s outside of your control, but there are positive things that have come about as a result of it. Unfortunately it’s dominating the work and focus.” As Ajzen’s theory of planned behavior (1991) asserted, the principals’ evaluations of teachers were influenced by their attitude toward the behavior, subjective norms of the school and district, and their perceived level of behavioral control. More specifically, principals’ intentions and behaviors were most notably shaped by their beliefs that the new rubric and growth scores would lead to favorable outcomes and improve teachers’ performance, the lack of support from district-
level leadership and lack of contact with site-level peers, and time constraints with limited
district-provided professional development targeted to the six dimensions.

Summary of Major Findings

There was a range of factors that influenced principals’ evaluations of teachers which
varied with individual principals and the building-specific contexts in which they worked. Not
surprisingly, when completing annual performance reviews principals in this case study favored
teachers who demonstrated strong teaching skills, were caring, ambitious, enthusiastic,
knowledgeable of their subject matter, worked well with others, and communicated effectively
(Murnane & Levy, 1996; Moss & Tilly, 1995; Harris, Rutledge, Ingle & Thompson 2006). As
discussed in Chapter Four, data on the perceived factors influencing principals’ evaluations of
teachers was collected, analyzed, and categorized into individualizing evaluations and
probationary versus non-probationary employment status (principal-teacher relationships),
workload, climate and culture, and parent perceptions and feedback (school level factors), and
the 5D+ evaluation rubric, district office support, isolation and lack of autonomy (district level
factors). A closer look at the interview records revealed principals broke down district office
support into the professional development, human resources (HR), cabinet (senior leadership),
instructional technology, and research and evaluation departments. The constant interaction of
data collection and analysis made it difficult to state in absolute terms how the principals
prioritized or sequenced the factors, but the major findings were: 1) principal-teacher
relationships within the specific school culture was the main determinant influencing the extent
to which principals engaged in critical evaluation of teachers, 2) the workload and job
complexity of building principals influenced the level of detail and diligence they dedicated to
their evaluation practice, 3) principals believed the overwhelming majority of PVSD teachers were effective or above and there was little they could do to improve or remove ineffective continuing contract teachers, 4) the new 5D+rubric was an improvement over the previous instrument and the professional development provided by the school district and University of Washington’s Center for Educational Leadership (CEL) enhanced principals’ efficacy in conducting teacher evaluations, and 5) district level leadership culture and support ultimately affected if, when, and how principals moved in the direction of creating improvement plans for teachers and/or pushing for transfer or dismissal.

Conclusions

Based on this study’s findings, it appears the premise of SB 191 that more frequent and intensive evaluations of teachers by principals will lead to higher levels of student learning is only valid if one of two conditions exist. The first is that teachers know how to improve student learning, but have not been sufficiently motivated to do so. The second is principals have the expertise to improve teachers’ professional practice through classroom observation, but have not done so. This effort to evaluate teachers to better performance is conceivably destined to fail if it keeps asking the wrong question. The question isn’t “How can I do a better job of monitoring teachers’ instructional performance,” but “How can we collectively do a better job monitoring students’ learning?” If SB 191 proves ineffective, the solution is not likely to be to reinforce an unproven and perhaps even misguided strategy demanding additional classroom observations, tighter supervision, and more punitive evaluations. Maybe what we need are “learning leaders” who create a school wide culture focused on learning for the students and adults who serve them. As Elmore (2004) stated, “The technical core of education… resides in individual classrooms,
not in the organizations that surround them” (pp. 5, 6). The management of schools then is essentially running the structures and processes around instruction, but not “managing” instruction.

Five conclusions consistent with the findings of Weisberg, et al. (2009) emerged from this study regarding principals’ perceptions of the influences on how they evaluate teachers:

1) Maintaining convivial social relations with teachers and a harmonious school culture appear to be stronger priorities for principals than enforcing strict interpretations of teacher performance evaluations.

2) New teacher performance evaluation processes may actually limit principals’ abilities to engage in “instructional leadership.”

3) Many of the deficiencies and limitations of the old teacher evaluation system appear to be present in the new system as well.

4) Formal in-service training on the understanding and use of the evaluation rubric, technology, and process through step-by-step instructions guiding principals’ analysis of classroom instruction using a common protocol appears to effect principals’ inter-rater reliability.

5) In general, principals conduct teacher performance evaluations primarily alone and with little or no opportunity for peer consultation.

**Conclusion #1 – Maintaining convivial social relations with teachers and a harmonious school culture appear to be stronger priorities for principals than enforcing strict interpretations of teacher performance evaluations.** Principals advocated for the importance of maintaining a healthy climate and culture grounded in caring relationships. The
underlying culture impacted principals’ perceptions regarding their evaluation decision making. The social pressures and subjective norm, or “the way we do things around here,” shaped principals’ beliefs. They recognized they were in “the people business,” and criticism was rarely constructive. As noted by Jacob (2011), some principals were ambivalent to confronting inept teaching out of fear it would negatively impact staff morale. One principal referenced his predecessor’s practice of renewing every teacher and changing his assistant principals’ summative ratings to effective if necessary to avoid conflict. He said, “We don’t hire to fire, we grow them, they’re ours.” Consistent with the old saying, “all politics is local,” principals realized the toll on their building culture that came with criticizing a tenured teacher’s performance. They felt it simply was not worth the fight. Principals acknowledged the importance of social capital, trust, and reciprocity as described by Bryk and Schneider (2002), and understood they were more likely to be fired from the bottom than the top, so if their teachers were happy then they were happy. The unintended, destructive consequences to a building’s climate and culture that resulted when a principal tried to coach a veteran out of the teaching profession could be disastrous. The teacher being criticized characteristically didn’t believe he/she was a bad teacher. He/she usually went to the union and made it personal. He/she didn’t see it as being about his/her practice. The principals reported when they went down this road it took a toll on their emotional psyche, especially if the teacher was a mainstay in the building, because they said horrible things about their administrators. Even if the most of the staff knew the teacher was doing bad things for kids, there was a culture of protection in the teaching ranks consistent with Siranni and Friedland’s (n. d.) proposition that people draw upon social trust, norms, and networks to make sense of situations and solve common problems. All of the principals had been directly or indirectly acquainted with adverse staff reactions to them doing their jobs, and it usually took just one incident for principals to pull back and focus on
teacher development over accountability. Principals did not see the 5D+ evaluation rubric as an accountability hammer; it was more about promoting teacher growth. The more experienced principals understood the faster and safer road to teacher dismissal was through progressive discipline. It was about compliance. Some principals went so far as to assert the 5D+ rubric should never have been turned into an evaluation tool, which is where the lines got blurred between professional development and evaluation. The rubric was created for professional development, and then PVSD made it about evaluation, which was not what the University of Washington intended. Principals wanted to be seen as coaches, but at the end of the day they were evaluators, and they had to use the same tool to coach as they did to evaluate. Principals wondered how prudent it was for them to use the professional development tool to prove a teacher was ineffective. Overall, principals were highly influenced by their relationships with their teachers and had a strong desire to maintain those relationships in a positive way. As this study illustrates, the school culture can neither be ignored nor discredited when explaining principals’ perceptions of their decision making.

**Conclusion #2 – New teacher performance evaluation processes may actually limit principals’ abilities to engage in “instructional leadership.”** The requirements of SB 191 added significant work to principals’ already full plates. Senior district leadership and policymakers ought to consider principal workload and job complexity when implementing new evaluation systems, as it appeared to be a significant deterrent to principals giving more thorough feedback to teachers and taking on performance concerns. Principals in this study reported one formal classroom observation took five to six hours to complete from initially walking into the classroom to delivering the written record after the post-observation conference, which was
consistent with DuFour and Mattos (2013) findings. The time it took to complete an observation was typically longer when the teacher had identified performance concerns. Furthermore, principals were tasked with annually evaluating their assistant principals, counselors, and licensed itinerant staff annually as well (i.e. school psychologists, occupational therapists, speech language pathologist, etc.). While the mandated requirement that every teacher be evaluated every year was well intentioned, the authors of SB 191 failed to recognize the crushing demands on the contemporary school principal. A synthesis of research by Marzano, Waters, and McNulty (2005) identified 21 different responsibilities that principals had to address in an environment where any or all of these responsibilities could suddenly be put on the back burner by crises over which the principal had little control. What principals prioritized their time doing said a great deal about the value they placed on various responsibilities (Pelz & Andrew, as cited in Taylor 1991), and the incentives to complete them (O’Reilly, 1983). Moreover, the Interstate School Leaders Licensure Consortium (ISLLC) Standards, model leadership standards that outlined what education leaders should know and be able to do to ensure all students who graduated from high school were prepared to enter college or the workforce, outlined foundational education leadership principles that cut across grade levels. The six standards, released in 1996 by the Council of Chief State School Officers, were expanded to fourteen standards and re-released in 2014. This was a clear indication of the intensifying workload and multidimensional role of school principals, of which teacher evaluation was just one part.

Principals said completing the sheer number of formal observations on every one of their licensed staff members every year would be challenging without strong support from their asst. principals and counselors. Principals felt they previously spent a lot of time in classrooms, but the new system pushed them to do more formal observations of thirty or more minutes each and
less informal five-ten minute walk-through observations, which they believed increased
reliability and were the most effective use of their time as noted in the Measures of Effective
Teaching (MET) project (2013). Since the adoption of the new rubric principals acknowledged it
took them about six weeks longer each semester to complete each round of the formal process on
their faculties, and as a result they had a lot less time to connect one-on-one with teachers, staff,
students, and parents.

Principal[s] offered suggestions to redistribute the workload. For instance, PVSD’s
Professional Practice Experts (PPEs), who served only as consultants to principals, could have
evaluated difficult-to-observe itinerant staff, such as, school psychiatrists, occupational
therapists, speech language pathologists, etc. Adding observations by observers like these from
outside the teacher’s school to those carried out by the teacher’s own administrator could provide
an ongoing check against in-school bias (MET Project, 2013). Another proposal was to allow
principals to assign teachers to a multi-year cycle, or revert back to the former three-year cycle
with those teachers who as one principal said, “We’re hitting the cover off the ball.” Principals
believed evaluating every teacher every year forced them to put the same amount of energy into
teachers who had been competent for a long time as teachers who had performance concerns.
Principal[s] also received a growth score on every teacher every year in the new system, which
served as a dip stick to tell a principal if something was going wrong. Principal[s] held that once a
teacher got to the level of a master teacher they were not likely to backslide to partially effective
or ineffective, and it wasn’t necessary to continue to evaluate them annually. Those teachers
would have to try really hard to fail at that point.
Conclusion #3 – Many of the deficiencies and limitations of the old teacher evaluation system appear to be present in the new system as well. Principals rate most teachers as effective or highly effective, and the perceived difficulty in removing ineffective or partially effective non-probationary teachers motivates principals to evaluate probationary teachers more critically. Principals’ single most often stated recommendation to benefit students and empower principals was to remove tenure. Without equivocation, principals affirmed that did not mean they wanted to fire everybody, it did not mean teachers could not disagree with them, and it did not mean they could not have a heated debate. Principals said that just knowing they had the authority to get rid of a non-probationary teacher would significantly raise the overall level of teaching in their buildings. As Hanushek (2010) asserted, simply replacing the bottom 6 to 10 percent of teachers with average teachers, not superstars, would raise student achievement levels in the United States from below average to near the top in comparison to other countries (2010).

Principals’ competitive notions of power (Kirby & Bogotch, 1993), and as a criteria in principals decision making (Pfeffer, 1992), tempered principals’ willingness to provide teachers with the necessary resources to improve teachers’ practice. As Callahan (1962) predicted, the passage of SB 191 attempted to improve teacher effectiveness via an evaluation system more analogous to corporate practice by changing how tenure was awarded and retained, yet the aggregate outcome of summative evaluations remained the same – almost all teachers were rated effective or highly effective, few were placed on improvement plans, and dismissal was not a likely course of action for non-probationary (continuing contract) teachers. It’s interesting that a statement like this could be made in a school district like PVSD, which received a high quantity and quality of teacher applications each spring relative to other districts in the state, especially at
the elementary level. It would be reasonable to assume principals could leverage the available teacher applicant pool and be more critical of their current faculty, especially if there were in fact a 100 or more teachers waiting in line for an opportunity to teach here. Yet, that was not the case. As noted by Adamowski, Therriault, and Cavanna (2007), the difference between the amount of authority principals thought they needed with regard to personnel decisions and the amount they actually had was a real barrier born from the negotiated agreement, district procedures, and SB 191. Principals speculated some of their teachers would claim their principals’ evaluations didn’t change much with the new system. In general, principals tried to differentiate their evaluations of teachers to get them to a better place through observing, listening, questioning, and supporting.

**Conclusion #4 – Formal in-service training on the understanding and use of the evaluation rubric, technology, and process through step-by-step instructions guiding principals’ analysis of classroom instruction using a common protocol appears to effect principals’ inter-rater reliability.** The principals overwhelmingly reported the 5D+ evaluation rubric’s six dimensions, sixteen sub-dimensions, and thirty-seven indicators were an improvement over the previous evaluation instrument’s six performance areas and twenty-seven criteria. More specifically, they preferred the 5D+ rubric’s marked band descriptors and the rating language of ineffective, partially effective, effective, and highly effective over the former unsatisfactory, needs improvement, and satisfactory. They believed the 5D+ rubric language brought a more systematic approach to conversations about effective instruction, while also acknowledging the enormous amount of information in it. As Taylor (1986) noted, making sense of information was a component of human behavior and critical to the decision making process.
Principals admittedly didn’t spend much time understanding what was behind the old evaluation system. They saw it as a checkbox they did once a year, without a common language or a way to talk directly to specific teaching behaviors. DeNisi and Kluger (2000) suggested effective feedback intended to motivate and improve performance should focus on the specific task and task performance, not on the individual person or the person’s self-concept. The feedback principals gave to teachers in the past was more or less their commentary about an area of instruction they thought the teacher was either good at or needed improvement in, and they could document in a formal observation write up. Conversely, principals believed the new 5D+ protocol incorporating noticings, wonderings, analysis, and feedback was rooted in sound research, best practices, and improved their self-efficacy (Bandura, 1989). It forced them to do more scripting, enhanced conversations about practice, helped principals ask better questions, identify specific areas to focus on, and shape the direction they wanted to go with teachers. Senge (1990) talked about this concept as the creative tension between current reality and perceived reality. The rubric encouraged releasing responsibility for learning to students by shifting the locus of control toward students doing the work with the teacher as the so-called guide on the side versus the more traditional sage on the stage. Principals likened their application of the 5D+ rubric to action research and collective inquiry. They felt the level of sophistication of the rubric helped satisfy the modern-day expectation that principals are instructional leaders, a role some contended was unrealistic and difficult to define. Still, most principals acknowledged approaching the evaluation system as an instructional coach, and had a positive experience with the 5D+ teaching and learning framework training as the foundation of the coaching they did with teachers. They thought it identified what effective teaching could look like, and it was the basis for rich, encouraging discussions with teachers. Principals said the hard
part was finding or constructing the supports and interventions to improve teachers’ instructional deficiencies.

**Conclusion #5 – In general, principals conduct teacher performance evaluations primarily alone and with little or no opportunity for peer consultation.** The limited support and sense of isolation from the district office reverberated throughout the study. This perception was based upon and reinforced by the subjective norms in the district. As a point of fact, five PVSD schools initially piloted the Colorado Dept. of Ed. evaluation rubric, ultimately adopted by nearly all 178 schools districts in the state, and all of a sudden PVSD principals were told the 5D+ was the instrument PVSD as going to use for evaluations. No vote was taken, it was just given to principals, even though the University of Washington hadn’t finished creating it and it wasn’t a proven rubric that had been used anywhere. Nonetheless, this coaching tool was reconfigured into an evaluation tool for the principals and their assistant principals to administer at the sites in a political climate that felt very accountability heavy. As Firestone and Gonzalez (2007) described, an accountability culture is top-down, reactive, and focused on increased efficiency, short term production, and using of data to hold individuals accountable rather than focusing on student learning, risk taking, and growing as an organization. Principals felt pressure from district office personnel to weed teachers out if they were perceived to not be doing the job, but rarely or never did those same district office staffers communicate to principals how to specifically address their biggest performance concerns.

While the principals were physically secluded in their own buildings, what they interpreted wasn’t in isolation; it was intrinsically connected to situational factors (Spillane & Miele, 2007). There was a perception that the district was looking at how principals evaluated
teachers so teachers could be moved on to greener pastures, and if they did not do that there was speculation the principal might be escorted out. This feeling was most palpable when principals recalled being given an index card at a district meeting, told to write down the names of their weakest teachers, and turn in the card. The teachers they wrote down were the ones the principals spent the most time helping improve. They were in those teachers classrooms more often and talked to them more often, but no one from the district walked with the principals through this process. Once it became clear a teacher was not right for a building, or the profession, few if any principals felt any inspiration to move toward dismissal. After all, if they provided coaching and critical feedback to a teacher on their own and experienced no interest from the district up to that point, why should they go down a path that would surely become more contentious, and risk being attacked? There simply was not a set of incentives (O’Reilly, 1983), or what Ajzen (1991) called perceived consequences that generate a favorable attitude toward doing a behavior, to encourage principals as decision makers to act authoritatively. Principals felt it safer and smarter to just wait the mediocre teacher out, provided he/she didn’t do anything too egregious. One principal even likened the behavior of principals in principals’ meetings to the behavior of a group of at-will probationary teachers who could be nonrenewed without cause.

Principals said the isolation inherent in conducting the evaluation process and one-on-one nature of evaluation-related meetings made performance review hard. Compounding this challenge was how teachers responded when principals brought their concerns to teachers’ attention and attempted to address deficiencies. Not surprisingly they said the attitudes and abilities of teachers made it easier; if teachers had a positive outlook, were open to feedback, wanted to improve, and took ownership. Winfield (1986) and Page (1991) identified teacher motivation and expectations for students as crucial factors for success with students. Teachers
who assumed personal and professional responsibility for students’ success and motivation, and maintained high standards were more likely to have students who succeeded. In contrast, teachers who believed students were responsible for their own motivation toward school, or some students were not capable of adequate performance were more likely to have students who were less successful (McLaughlin & Talbert, 1993; Knapp, 1995).

Typically the second phase in the evaluation process was to place the teacher on an improvement plan, but if the teacher didn’t meet the expected outcomes of the improvement plan, he or she was placed on another improvement plan the next semester or year. There was ambiguity from the district about how inadequate a teacher’s performance had to be before placing that teacher on an improvement plan, and how much that ineffectiveness needed to be documented before going to the third phase of the evaluation process; dismissal. Most principals resigned themselves to the fact that these processes took years, required district approval, and were not time well spent. Of course, principals understood if a teacher committed a felony this process moved quicker. They commented the evaluation process, even when administered correctly, would only get a marginal teacher marked partially effective at best. The more experienced principals knew a letter of directive had more punch than the phases of the evaluation process and an improvement plan. If a teacher violated district policy he/she would likely receive a letter of directive. A violation of the letter of directive could lead to insubordination, which could then lead to neglect of duty and dismissal.

Principals said they grew accustomed to evaluating teachers on their own with little input from supervisors or colleagues outside their buildings. This practice was contrary to the recommendations of the MET Project (2013) which advocated adding a second observer to increase reliability, conducting additional shorter observations, and adding observations by
observers outside the teacher’s school. Since the high stakes penalties of SB 191, like the loss of non-probationary status after two years of consecutive low ratings had yet to go into effect, principals generally conducted business as usual. They discussed teacher performance only with their assistant principals, and some principals delighted in the absence of outside involvement, as in lessened the likelihood their personnel decisions would be micro-managed by the district.

Fitzpatrick (1989) reviewed research from social psychology on the effects of group discussion and exposure to other group member’s opinions. She found that group discussion may perhaps polarize preexisting opinions or cause “social comparison,” with subjective judgments the most susceptible to this influence, which is consistent with Ajzen’s (1991) notion of normative beliefs.

Reflecting on the spirit and accountability components of SB 191, and whether it would change the way they evaluated teachers over time, most principals confirmed there has been no attack on tenure. If there had been, non-probationary status would have been removed by now, as the original legislation was designed to take effect during the 2012-13 year. Still, most principals expressed gratitude that SB 191 finally put some teeth into the evaluation process. They shared stories about teachers who received satisfactory evaluations for decades, but who they would describe as some of the worst teachers they had ever seen, and no one ever said anything. They talked about principals being afraid, because they were supervising and evaluating teachers who were protected by tenure and principals were not. They felt teachers could get rid of principals quicker than they could get rid of teachers. Principals hoped the new evaluation process would make them less fearful. The interviews revealed what many principals observed in their colleagues or were guilty of themselves, looking the other way when performance issues were staring them right in the face. While images like this possibly have long been suspected in public education they have rarely been exposed, unless a principal departed and a new one took over the
Building, largely because of the isolated and confidential nature of principals’ evaluations of teachers.

**Recommendations**

Understanding the social forces influencing principals’ decision-making processes and whether or not and how they pursue teacher renewal, remediation, or dismissal is a necessary complement to the ongoing scholarly debate surrounding the larger issue of teacher evaluation practice in the complex environment of public schools considering the increased scrutiny of principals to improve student experiences and outcomes. This study contributed to this line of inquiry. I conclude with recommendations for supporting building principals and improving the delivery of teacher evaluation systems.

1) Provide ongoing *professional development* on evaluation documents, technology, and processes differentiated to principals’ needs. Include a menu of comments under each evaluation indicator similar to how teachers add comments to student report cards. Align all professional development offerings for teachers with the local evaluation rubric.

2) Foster *trust* through openness between principal and teacher employee groups by presuming good intentions and speaking from the heart and mind to create a culture that honors individual dignity and respect.

3) Create a *collaborative evaluation rounds protocol*, through neighborhood and/or programmatic feeder schools, allowing principals to conduct teacher evaluations in partnership with site-level colleagues and/or district-level leaders, and add a second observer to increase reliability and provide a check against in-school bias.
4) Eliminate non-probationary (tenure) status.

5) Increase principal staffing with administrative assistant support at school sites to lessen the workload and enhance the delivery of a more detailed evaluation system.

Train office managers on how to support principals as executive assistants, beyond their more traditional roles as secretaries and bookkeepers.

Principals ought not to be saddled with administering new teacher evaluation processes alone. The myth of the heroic principal triumphing against all odds undermines the importance of cultural and structural supports. Building a strong professional community and maintaining a conversation within districts and between schools about effective evaluation practice may provide the necessary supports and resources necessary for principals make important evaluation decisions.
References


Implications for values and leadership. *Journal of Educational Administration, 42*(2), 132-136.


Калхан, Р. (1962). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools*. Chicago: University of Chicago Press.


DuFour, R., & Mattos, M. (2013, April). How Do Principals Really Improve Schools?


Goldhaber, D. (2002, Spring). Mystery of Good Teaching: The evidence shows that good teachers make a clear difference in student achievement. The problem is that we don't
really know what makes a good teacher. *Education Next*. Hoover Institution Press.


National Council on Teacher Quality. (2009, August). *Race to the Top: Colorado may be used to high altitudes, but can it compete in Race to the Top?* Retrieved from http://www.nctq.org/p/publications/docs/NCTQ_CO_Race_to_the_Top.pdf


Rimmer, J. (2011, December). *5Ds Instructional Leadership Assessment Results*. Presentation by the University of Washington Center for Educational Leadership to Pioneer Valley School District administrators (http://www.k-12leadership.org/faq#slideshow). Fort Collins, CO.


Rowan, B., Correnti, R., & Miller, R. J. (2002). *What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools*. Teachers College Record, 104, 1,525-1,567.


Appendix A: Participation Solicitation Cover Letter

(Date)

(Recipient Name)
(Title)
(School Name)
(Street Address)
(City, ST, Zip)

Dear (Recipient Name),

As you may know, in addition to my role as a middle school principal in the Pioneer Valley School District (PVSD), I’m also a doctoral candidate at Pennsylvania State University. My dissertation title is *Influences on Principal Evaluations of Teacher Performance*. Using a case study design, and Ajzen’s theory of planned behavior as the conceptual framework, the purpose of my research is to investigate how PVSD principals’ attitudes toward a behavior (behavioral beliefs), social pressure (normative beliefs), and behavioral controls (control beliefs) influence their decisions when evaluating teacher performance within the school, district, and state contexts of conflicting social forces and collegial interactions per Colorado Senate Bill 191.

I’m writing today to ask for your voluntary participation in this study. My goal is to interview 15 of PVSD’s 45 building principals. Your commitment would involve completing a 45-minute one-on-one interview with me. If you are willing to be involved, please reply to this email to confirm your participation, and I will then work with you to schedule the one-on-one interview in the next 4 weeks. I’ve attached the Signed Consent Form to Participate in a Social Science Research Study, which has more specific information. Of course, your contributions to this study will be kept confidential.

I would sincerely appreciate your help in completing this study. I realize the demands on your time as a building principal. Your involvement will assist me in understanding PVSD principals’ perceptions, interpretations, and actions in the context of evaluating teachers, and you will gain personal insight into your own behavioral intentions. If you have any questions please contact me via email: tjd188@psu.edu, or cell: 970-567-2410. Thank you in advance for your support.

Sincerely,

Tom Dodd
PhD Candidate (ABD)
PSU Educational Leadership
Cell: 970-567-2410
Email: tjd188@psu.edu

Attached: Signed Consent Form to Participate in a Social Science Research Study
Appendix B: Signed Consent Form to Participate in a Social Science Research Study  
The Pennsylvania State University Office for Research Protections (ORP)

Project Title: Influences on Principal Evaluations of Teacher Performance  
Principal Investigator: Tom Dodd (970-567-2410, tjd188@psu.edu)  
Dissertation Advisor: Dr. Nona Prestine (814-863-3762, nap11@psu.edu)  
Department: Education Policy Studies

You were selected as a possible participant for this voluntary research study within the Pioneer Valley School District (PVSD) of Colorado because you currently serve as a PVSD principal and evaluate teachers annually. This form provides information about this study. Please read this form and ask any questions you may have before agreeing to take part in this study.

Purpose of this Research:  
The purpose of this study is to investigate how principals’ attitudes toward a behavior (behavioral beliefs), social pressure (normative beliefs), and behavioral controls (control beliefs) influence principals’ decisions when evaluating teacher performance within the school, district, and state contexts of conflicting social forces and the collegial interactions.

Number of Participants:  
Fifteen of PVSD’s forty-five K-12 principals (minus the lead investigator) will take part in this study.

Procedures to be Followed:  
If you agree to participate in this study you will be asked to:
1) Complete a one-on-one interview in person with the principal investigator. Participants are free to skip any questions they prefer not to answer. The interview will be conducted at a location convenient for you and will be audiotaped. Interviews will take approximately 45 minutes to complete.
2) Provide, at your discretion, any documents you believe may be useful to this study. All participants will be provided with an electronic copy of the study at its conclusion.

This Study has the Following Risks:  
The written audiotaped interview questions are designed for participants to provide perspectives and opinions about their experiences conducting teacher evaluations. If the participant believes it necessary to provide information about a specific individual, group of people, or the school district to answer a question, and that information is considered negative in nature, the following precautions will be taken to reduce the possibility of risk and discomfort from participating in their interview:
1) Participants will be free to discontinue the interview at any time.
2) The researcher will not name participants or their specific school or the district in any research reports stemming from this study.
3) Interview audiotapes will be played only for transcription and coding, stored in a locked cabinet in the principal investigator’s office, and destroyed at the conclusion of the study.
Benefits of Participation in this Study:
Benefits include participating in a study affiliated with the Pennsylvania State University College of Education, and making an intellectual contribution to building an analytical hypothesis regarding the behavior of principals in assessing teacher performance that holds across school, district, and state contexts and can be refined and tested in future research. Your cooperation is appreciated by doctoral candidate Tom Dodd, who is conducting this study.

Voluntary Participation in this Study:
Data collection will take place during the fall semester of the 2014-15 school year, and participation is voluntary. If you decide to participate, you may decline to answer any question and you may choose to withdraw at any time. If you decline participation or withdraw, you will not be penalized or lose benefits or services unrelated to this study.

Statement of Confidentiality:
The PSU Office for Research Protections and the Social Science Institutional Review Board (IRB) may inspect and/or copy your research records for quality assurance and data analysis related to this project. All research data will be kept confidential and stored securely. Only the approved researchers (principal investigator and dissertation advisor) will have access to the research records. PVSD and participants will be given pseudonyms to protect identities. In the event of any publication or presentation resulting from this research, no personally identifiable information will be shared without your permission.

Compensation:
You will not be compensated for you time, or participation in this study.

Contacts and Questions:
If you have questions or concerns about this research project, please contact the researcher(s):
Principal Investigator: Tom Dodd (970-567-2410, tjd188@psu.edu)
Dissertation Advisor: Dr. Nona Prestine (814-863-3762, nap11@psu.edu)

If you have questions about your rights as a research participant, or concerns about the research and wish to talk to someone other than the research team, or if you cannot reach the research team, you may contact The Pennsylvania State University Office for Research Protections for Human Participants Research at 814-865-1775, or ORProtections@psu.edu.

Statement of Informed Consent:
I have read the above information, asked questions, and received satisfactory answers. I consent to participate in this study, and will receive a copy of this consent form for my records.

<table>
<thead>
<tr>
<th>Participant Signature</th>
<th>Printed Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Person Obtaining Consent</th>
<th>Printed Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C: Interview Questions to Elicit Commonly Held Beliefs
Using a Conversational Approach

NAME: ________________________ PVSD ID #: _______________ DATE: ______________

Opening
As you know, our school district has a new teacher evaluation process and rubric. As building principals, we are the frontline implementers of this new approach.

(1) Tell me about your experiences with the new evaluation process?

(2) What do you see as the favorable or unfavorable aspects of evaluating teachers using the new rubric?

(3) Are there any individuals, or groups of people you are especially mindful of when writing evaluations?

(4) What factors or circumstances make evaluating teachers easy or difficult?

Potential Follow Up Probing Questions
You mentioned thus and so, what did you mean by that?
It’s interesting you mentioned thus and so, can you give me an example of that?
Section 1: Please answer these demographic questions about yourself.
A. Gender: MALE _____ FEMALE _____
B. Age: __________
C. Ethnicity: ______________________________
D. Highest degree attained: ______________________________
E. How many years have you been a principal? __________
F. How many years have you been the principal of your current school? __________
G. How many teachers will be evaluated in your school this year? __________
H. How many of those teachers will you directly evaluate this year? __________
I. Do you have an assistant principal(s)? YES _____ NO _____
J. If yes, how many teachers will your assistant principal(s) directly evaluate this year? _____
K. Have you ever personally received a less than effective overall performance evaluation? YES _____ NO _____

Section 2: Please answer these demographic questions about your school.
A. School: _______________________________________________________
B. Elementary: _____ Middle: _____ High: _____ K-12: _____ Other: _____
C. Enrollment: __________
D. Percent of students eligible for free/reduced lunch support: __________
E. Percent of students who identify with a minority group: __________
F. Percent of students with an Individualized Education Plan (IEP): __________
G. Percent of students with a gifted and talented identification: __________
H. 2013-14 Colorado Dept. of Ed. School Performance Framework (SPF) rating:
   3-year: ____________________ 1-year: ____________________
Appendix E: Glossary

| **1st Strata** | The use of district-wide assessments that have been scored externally within a statistical process that yields teacher-effect estimates useful in determining Growth Ratings related to SB-191. |
| **2nd Strata** | The use of district-wide assessments that have been scored internally within a statistical process that yields teacher-effect estimates useful in determining Growth Ratings related to SB-191. |
| **3rd Strata** | The system by which PVSD seeks evidence of effectiveness in the form of student outcomes when Growth Ratings of a 3 or 4 are not available via the 1st or 2nd Strata processes. The Student Learning Objectives (SLO) process is utilized. |
| **Criterion Referenced** | A criterion-referenced teacher effectiveness measure indicates whether the teacher's student-outcomes are higher or lower than some criteria that are not dependent on other teachers’ student-outcomes. |
| **Growth Rating Process** | The process by which PVSD will determine individual educator's Growth Ratings as required by SB-191. |
| **Norm Referenced** | A norm-referenced teacher effectiveness measure indicates whether the teacher's student outcomes are higher or lower than other teachers who were able to be included in the norming group. |
| **SB 191** | Colorado Senate Bill 191, otherwise known as Educator Effectiveness. |
| **Student Learning Objective (SLO)** | A participatory method of setting measurable goals, or objectives for a specific assignment or class, in a manner aligned with the subject matter taught, and in a manner that allows for the evaluation of the baseline performance of students and the measurable gains in student performance during the course of instruction. |
| **Student Growth** | Measurable student learning over time. |
| **Success Criteria** | The minimum score on a summative assessment that indicates student proficiency regarding the standards being assessed. |
| **Teacher’s Objective** | An individualized performance target that earns the Growth Rating of “Expected” and is based on the percentage of students meeting the student success criteria defined explicitly in the SLO form. |
Appendix F: PVSD Teacher Evaluation Process Overview
Appendix G: PVSD-designed Value-added Measure (VAM) – A 3-tiered Stratified Approach to Identifying Evidence of Growth

Apply the rating with the highest reliability demonstrating expected levels of student growth.

1st Strata: HLM Growth Modeling
Unconditional Growth Estimates - No statistical adjustments to expectations
Conditional Growth Estimates - Adjustments for important student factors
  • Student IEP status (State Required), Student attendance rate below 95%, Student starting in the top quartile

2nd Strata: HLM Student Learning Measures
  • Used when evidence of expected growth not available from 1st Strata
  • Unconditional and conditional growth estimates calculated as in 1st Strata, but using local measures

3rd Strata: Student Learning Objectives (SLOs): Participatory Goal Setting with Principal
  • Used when evidence of expected growth not available from 1st or 2nd Strata
  • All new hires will receive their Growth Rating as a component of a revised Induction Process
  • Alternate Ed. Campus Staff may have non-academic SLOs due to 95% “High Risk” students
  • Integrated Services teachers may demonstrate progress toward functional goals in IEPs
Appendix H: PVSD Shared Educator Growth Rating Calculation

Turnaround = 1, Priority Improvement = 2, Improvement = 3, Performance = 4

Shared Measure of Growth = Final Educator Growth Rating \cdot 0.80 (Individual Growth Rating) + 0.20 (Shared Growth Rating). The most recent School Performance Framework Plan Assignment from the Colorado Dept. of Ed. will be used as the Shared Growth Rating.
Appendix I: PVSD Teacher Rating Calculation

### Combining Professional Practice & Growth

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Combined Rating:**
- 4 = Highly Effective
- 3 = Effective
- 2 = Partially Effective
- 1 = Ineffective

**Prof. Practice Rating:**
- 4 = Highly Effective
- 3 = Effective
- 2 = Partially Effective
- 1 = Ineffective

**Expected Growth Range:**
- Effect Size $0 \pm 0.25$
- $\approx \mu \pm 1\sigma$ of 1st Strata Teachers' Effects
- $\approx$ Middle 68%
- 60% - 80% of students meet Learning Outcome Targets

**Growth Rating:**
- 4 = High Growth
- 3 = Expected Growth Range
- 2 = Low Growth
- 1 = Significant Low Growth
## 5D+™ Teacher Evaluation Rubric

We know that building the capacity of teachers will lead to better instruction and greater learning for all students. Helping educators understand what good teaching looks like is at the heart of the Center for Educational Leadership’s 5D+ Teacher Evaluation Rubric—a growth-oriented tool for improving instruction.

### Dimensions of the 5D+ Teacher Evaluation Rubric

The 5D+ Teacher Evaluation Rubric is based on the 5 Dimensions of Teaching and Learning™ (5D™) instructional framework, which is derived from an extensive study of research on the core elements that constitute quality instruction. These core elements have been incorporated into the 5D framework and 5D+ rubric as five dimensions—Purpose, Student Engagement, Curriculum & Pedagogy, Assessment for Student Learning, and Classroom Environment & Culture—which are divided into 13 subdimensions. The 5D+ rubric also includes Professional Collaboration and Communication, which is based on activities and relationships that teachers engage in outside of classroom instruction.

### Performance Levels

Performance levels within each indicator are used to delineate teaching practice, from ineffective to partially effective, effective and highly effective. The sophistication of teaching practice and the role of students increase across the levels of performance. The language describing each performance level has been carefully examined by a psychometrician to assure clarity, to avoid the risk of a teacher being rated more than once for similar teaching behavior, and to ensure that each indicator evaluates only one aspect of teaching practice. A careful analysis of instructional practice leads to the determination of a teacher’s performance level on each indicator.

### Organization of the 5D+ Teacher Evaluation Rubric

The 5D+ rubric is composed of 37 indicators of teacher performance, each appearing on a separate page of the rubric. In the example below: the dimension is Purpose, the subdimension is Standards, and the indicator is Connection to Standards, Broader Purpose and Transferable Skill. The pages are color-coded by dimension.

### Colorado Performance Evaluation Ratings

The titles of the performance levels on this document—“ineffective,” "partially effective," "effective" and "highly effective"—match the performance evaluation ratings for teachers required by the Colorado Department of Education.

### Resources and Support

The 5D+ Teacher Evaluation Rubric is also available, with associated resources, on the University of Washington Center for Educational Leadership (CEL) website at www.k-12leadership.org/teacher-eval. The performance level titles used on the posted rubric are: "unsatisfactory," "basic," "proficient," and "distinguished."
## 5D+™ Teacher Evaluation Rubric

### Purpose

<table>
<thead>
<tr>
<th>Ineffective</th>
<th>Partially Effective</th>
<th>Effective</th>
<th>Highly Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Standards: Connection to standards, broader purpose and transferable skill

**P1**

- Ineffective: The lesson is not based on grade level standards. There are no learning targets aligned to the standard. The lesson does not link to broader purpose or a transferable skill.
- Partially Effective: The lesson is based on grade level standards and the learning target(s) align to the standard. The lesson is occasionally linked to broader purpose or a transferable skill.
- Effective: The lesson is based on grade level standards and the learning target(s) align to the standard. The lesson is frequently linked to broader purpose or a transferable skill.
- Highly Effective: The lesson is based on grade level standards and the learning target(s) align to the standard. The lesson is consistently linked to broader purpose or a transferable skill.

### Standards: Connection to previous and future lessons

**P2**

- Ineffective: The lesson is rarely or never linked to previous and future lessons.
- Partially Effective: The lesson is clearly linked to previous and future lessons.
- Effective: The lesson is clearly linked to previous and future lessons. Lessons build on each other in a logical progression.
- Highly Effective: The lesson is clearly linked to previous and future lessons. Lessons build on each other in ways that enhance student learning. Students understand how the lesson relates to previous lesson.

### Teaching Point: Teaching point(s) are based on students’ learning needs

**P3**

- Ineffective: Teacher rarely or never bases the teaching point(s) on students’ learning needs – academic background, life experiences, culture and language.
- Partially Effective: Teacher bases the teaching point(s) on limited aspects of students’ learning needs – academic background, life experiences, culture and language.
- Effective: Teacher bases the teaching point(s) on the learning needs – academic background, life experiences, culture and language – for some groups of students and individual students.
- Highly Effective: Teacher bases the teaching point(s) on the learning needs – academic background, life experiences, culture and language – for some groups of students and individual students.

### Learning Target: Communication of learning target(s)

**P4**

- Ineffective: Teacher rarely or never states or communicates with students about the learning target(s).
- Partially Effective: Teacher states the learning target(s) at the beginning of each lesson.
- Effective: Teacher communicates the learning target(s) through verbal and visual strategies, checks for student understanding of what the target(s) are.
- Highly Effective: Teacher communicates the learning target(s) through verbal and visual strategies, checks for student understanding of what the target(s) are and references the target throughout instruction.

### Learning Target: Success criteria and performance task(s)

**P5**

- Ineffective: The success criteria for the learning target(s) are nonexistent or aren’t clear to students.
- Partially Effective: The success criteria for the learning target(s) are clear to students. The performance tasks align to the success criteria in a limited manner.
- Effective: The success criteria for the learning target(s) are clear to students. The performance tasks align to the success criteria.
- Highly Effective: The success criteria for the learning target(s) are clear to students. Students refer to success criteria and use them for improvement.
## 5D+™ Teacher Evaluation Rubric

<table>
<thead>
<tr>
<th>Student Engagement</th>
<th>Ineffective</th>
<th>Partially Effective</th>
<th>Effective</th>
<th>Highly Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SE1</strong> Intellectual Work: Quality of questioning</td>
<td>Teacher rarely or never asks questions to probe and deepen students' understanding or uncover misconceptions.</td>
<td>Teacher occasionally asks questions to probe and deepen students' understanding or uncover misconceptions.</td>
<td>Teacher frequently asks questions to probe and deepen students' understanding or uncover misconceptions. Teacher assists students in clarifying their thinking with one another.</td>
<td>Teacher frequently asks questions to probe and deepen students' understanding or uncover misconceptions. Teacher assists students in clarifying their thinking with one another. Students question one another to probe for deeper thinking.</td>
</tr>
<tr>
<td><strong>SE2</strong> Intellectual Work: Ownership of learning</td>
<td>Teacher rarely or never provides opportunities and strategies for students to take ownership of their own learning to develop, test and refine their thinking.</td>
<td>Teacher occasionally provides opportunities and strategies for students to take ownership of their learning. Locus of control is with teacher.</td>
<td>Teacher provides opportunities and strategies for students to take ownership of their learning. Some locus of control is with students in ways that support students' learning.</td>
<td>Teacher consistently provides opportunities and strategies for students to take ownership of their learning. Most locus of control is with students in ways that support students' learning.</td>
</tr>
<tr>
<td><strong>SE3</strong> Engagement Strategies: High cognitive demand</td>
<td>Teacher expectations and strategies engage few or no students in work of high cognitive demand.</td>
<td>Teacher expectations and strategies engage some students in work of high cognitive demand.</td>
<td>Teacher expectations and strategies engage most students in work of high cognitive demand.</td>
<td>Teacher expectations and strategies engage all students in work of high cognitive demand.</td>
</tr>
<tr>
<td><strong>SE4</strong> Engagement Strategies: Strategies that capitalize on learning needs of students</td>
<td>Teacher rarely or never uses strategies based on the learning needs of students – academic background, life experiences, culture and language of students.</td>
<td>Teacher uses strategies that capitalize and are based on learning needs of students – academic background, life experience and culture and language of students – for the whole group.</td>
<td>Teacher uses strategies that capitalize and are based on learning needs of students – academic background, life experiences, culture and language of students – for the whole group and small groups of students.</td>
<td>Teacher uses strategies that capitalize and build upon learning needs of students – academic background, life experiences, culture and language of students – for the whole group and small groups of students and individual students.</td>
</tr>
<tr>
<td><strong>SE5</strong> Engagement Strategies: Expectation, support and opportunity for participation and meaning making</td>
<td>Teacher rarely or never uses engagement strategies and structures that facilitate participation and meaning making by all students. Few students have the opportunity to engage in quality talk.</td>
<td>Teacher uses engagement strategies and structures that facilitate participation and meaning making by students. Some students have the opportunity to engage in quality talk.</td>
<td>Teacher sets expectation and provides support for a variety of engagement strategies and structures that facilitate participation and meaning making by students. Most students have the opportunity to engage in quality talk.</td>
<td>Teacher sets expectation and provides support for a variety of engagement strategies and structures that facilitate participation and meaning making by students. All students have the opportunity to engage in quality talk. Routines are often student-led.</td>
</tr>
<tr>
<td><strong>SE6</strong> Talk: Substance of student talk</td>
<td>Student talk is nonexistent or is unrelated to content or is limited to single-word responses or incomplete sentences directed to teacher.</td>
<td>Student talk is directed to teacher. Talk associated with content occurs between students, but students do not provide evidence for their thinking.</td>
<td>Student-to-student talk reflects knowledge and ways of thinking associated with the content. Students provide evidence to support their thinking.</td>
<td>Student-to-student talk reflects knowledge and ways of thinking associated with the content. Students provide evidence to support their arguments and new ideas.</td>
</tr>
</tbody>
</table>
### Curriculum & Pedagogy

<table>
<thead>
<tr>
<th></th>
<th>Ineffective</th>
<th>Partially Effective</th>
<th>Effective</th>
<th>Highly Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CP1</strong> Curriculum: Alignment of instructional materials and tasks</td>
<td>Instructional materials and tasks rarely or never align with the purpose of the unit and lesson.</td>
<td>Instructional materials and tasks align with the purpose of the unit and lesson. Materials and tasks frequently align with student's level of challenge.</td>
<td>Instructional materials and tasks align with the purpose of the unit and lesson. Materials and tasks consistently align with student's level of challenge.</td>
<td></td>
</tr>
<tr>
<td><strong>CP2</strong> Teaching Approaches and/or Strategies: Discipline-specific conceptual understanding</td>
<td>Teacher rarely or never uses discipline-specific teaching approaches and strategies that develop students' conceptual understanding.</td>
<td>Teacher occasionally uses discipline-specific teaching approaches and strategies that develop students' conceptual understanding.</td>
<td>Teacher frequently uses discipline-specific teaching approaches and strategies that develop students' conceptual understanding.</td>
<td>Teacher consistently uses discipline-specific teaching approaches and strategies that develop students' conceptual understanding.</td>
</tr>
<tr>
<td><strong>CP3</strong> Teaching Approaches and/or Strategies: Pedagogical content knowledge</td>
<td>Instruction is rarely or never consistent with pedagogical content knowledge and does not support students in discipline-specific habits of thinking.</td>
<td>Instruction is occasionally consistent with pedagogical content knowledge and supports students in discipline-specific habits of thinking.</td>
<td>Instruction is frequently consistent with pedagogical content knowledge and supports students in discipline-specific habits of thinking.</td>
<td>Instruction is always consistent with pedagogical content knowledge and supports students in discipline-specific habits of thinking.</td>
</tr>
<tr>
<td><strong>CP4</strong> Teaching Approaches and/or Strategies: Teacher knowledge of content</td>
<td>Teacher demonstrates a lack of knowledge of discipline-based concepts by making content errors.</td>
<td>Teacher demonstrates a basic knowledge of how discipline-based concepts relate to or build upon one another.</td>
<td>Teacher demonstrates a solid understanding of how discipline-based concepts relate to or build upon one another. Teacher identifies and addresses student misconceptions in the lesson or unit.</td>
<td>Teacher demonstrates an in-depth understanding of how discipline-based concepts relate to or build upon one another. Teacher identifies and addresses student misconceptions that impact conceptual understanding over time.</td>
</tr>
<tr>
<td><strong>CP5</strong> Teaching Approaches and/or Strategies: Differentiated instruction</td>
<td>Teacher rarely or never uses strategies that differentiate for individual learning strengths and needs.</td>
<td>Teacher occasionally uses strategies that differentiate for individual learning strengths and needs.</td>
<td>Teacher frequently uses strategies that differentiate for individual learning strengths and needs.</td>
<td>Teacher consistently uses strategies that differentiate for individual learning strengths and needs.</td>
</tr>
<tr>
<td><strong>CP6</strong> Scaffolds for Learning: Scaffolds the task</td>
<td>Teacher provides limited scaffolds and structures that may or may not be related to and support the development of the targeted concepts and/or skills.</td>
<td>Teacher provides scaffolds and structures that are clearly related to and support the development of the targeted concepts and/or skills.</td>
<td>Teacher provides scaffolds and structures that are clearly related to and support the development of the targeted concepts and/or skills.</td>
<td>Teacher provides scaffolds and structures that are clearly related to and support the development of the targeted concepts and/or skills. Students use scaffolds across tasks with similar demands.</td>
</tr>
<tr>
<td><strong>CP7</strong> Scaffolds for Learning: Gradual release of responsibility</td>
<td>Teacher rarely or never uses strategies for the purpose of gradually releasing responsibility to students to promote learning and independence.</td>
<td>Teacher occasionally uses strategies for the purpose of gradually releasing responsibility to students to promote learning and independence.</td>
<td>Teacher frequently uses strategies for the purpose of gradually releasing responsibility to students to promote learning and independence.</td>
<td>Teacher consistently uses strategies for the purpose of gradually releasing responsibility to students to promote learning and independence. Students expect to be self-reliant.</td>
</tr>
</tbody>
</table>
### 5D+™ Teacher Evaluation Rubric

<table>
<thead>
<tr>
<th>Assessment for Student Learning</th>
<th>Ineffective</th>
<th>Partially Effective</th>
<th>Effective</th>
<th>Highly Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assessment: Self-assessment of learning connected to the success criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1 Students are rarely or never given an opportunity to assess their own learning in relation to the success criteria for the learning target.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2 Assessment: Demonstration of learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3 Assessment: Formative assessment opportunities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4 Assessment: Collection systems for formative assessment data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5 Assessment: Student use of assessment data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6 Adjustments: Teacher use of formative assessment data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Version 2 / Colorado**

COPYRIGHT ©2012 UNIVERSITY OF WASHINGTON, CENTER FOR EDUCATIONAL LEADERSHIP
# 5D+™ Teacher Evaluation Rubric

## Classroom Environment & Culture

<table>
<thead>
<tr>
<th>CEC1</th>
<th>Use of Physical Environment: Arrangement of classroom</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
<td>Physical environment of the room is unsafe and the arrangement gets in the way or distracts from student learning and the purpose of the lesson.</td>
</tr>
<tr>
<td><strong>Partially Effective</strong></td>
<td>The physical environment is safe but the arrangement neither supports nor distracts from student learning or the purpose of the lesson.</td>
</tr>
<tr>
<td><strong>Effective</strong></td>
<td>The physical environment is safe, and the arrangement supports student learning and the purpose of the lesson.</td>
</tr>
<tr>
<td><strong>Highly Effective</strong></td>
<td>The physical environment is safe, and the arrangement supports student learning and the purpose of the lesson. Teacher and students use the physical arrangement for learning.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CEC2</th>
<th>Use of Physical Environment: Accessibility and use of materials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
<td>The resources, materials and technology in the classroom do not relate to the content or current units studied, or are not accessible to all students to support their learning during the lesson.</td>
</tr>
<tr>
<td><strong>Partially Effective</strong></td>
<td>The resources, materials and technology in the classroom relate to the content or current unit studied and are accessible to all students but are not referenced by teacher.</td>
</tr>
<tr>
<td><strong>Effective</strong></td>
<td>The resources, materials and technology in the classroom relate to the content or current unit studied, are accessible to all students and are intentionally used by teacher to support learning.</td>
</tr>
<tr>
<td><strong>Highly Effective</strong></td>
<td>The resources, materials and technology in the classroom relate to the content or current unit studied, are accessible to all students and are intentionally used by both teacher and student to support learning. Students are familiar and comfortable with using the available resources.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CEC3</th>
<th>Classroom Routines and Rituals: Discussion, collaboration and accountability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
<td>Routines for discussion and collaborative work are absent, poorly executed or do not hold students accountable for their work and learning.</td>
</tr>
<tr>
<td><strong>Partially Effective</strong></td>
<td>Routines for discussion and collaborative work are present, but may not result in effective discourse. Students are held accountable for completing their work but not for learning.</td>
</tr>
<tr>
<td><strong>Effective</strong></td>
<td>Routines for discussion and collaborative work have been taught, are evident, and result in effective discourse related to the lesson purpose. Students independently use these routines during the lesson. Students are held accountable for their work and learning.</td>
</tr>
<tr>
<td><strong>Highly Effective</strong></td>
<td>Routines for discussion and collaborative work have been explicitly taught, are evident, and result in effective discourse related to the lesson purpose. Students independently use the routines during the lesson. Students are held accountable for their work, take ownership for their learning and support the learning of others.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CEC4</th>
<th>Classroom Routines and Rituals: Use of learning time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
<td>Teacher or students frequently disrupt or interrupt learning activities, which results in loss of learning time. Transitions are disorganized and result in loss of instructional time.</td>
</tr>
<tr>
<td><strong>Partially Effective</strong></td>
<td>Teacher or students occasionally disrupt or interrupt learning activities, which results in some loss of learning time. Some transitions are disorganized and result in loss of instructional time.</td>
</tr>
<tr>
<td><strong>Effective</strong></td>
<td>Learning time is mostly maximized in service of learning. Transitions are teacher-dependent and maximize instructional time.</td>
</tr>
<tr>
<td><strong>Highly Effective</strong></td>
<td>All available time is maximized in service of learning. Student behavior is appropriate. Students manage themselves and each other in managing behavior, or there is no student misbehavior.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CEC5</th>
<th>Classroom Routines &amp; Rituals: Managing student behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
<td>Teacher rarely or never responds to student misbehavior by following classroom routines and/or building discipline procedures. Student behavior does not change or may escalate.</td>
</tr>
<tr>
<td><strong>Partially Effective</strong></td>
<td>Teacher responds to student misbehavior by following classroom routines and building discipline procedures, but with uneven student behavior results.</td>
</tr>
<tr>
<td><strong>Effective</strong></td>
<td>Teacher responds to student misbehavior by following classroom routines and building discipline procedures. Student misbehavior is rare.</td>
</tr>
<tr>
<td><strong>Highly Effective</strong></td>
<td>Teacher responds to student misbehavior by following classroom routines and building discipline procedures. Student behavior is appropriate. Students manage themselves, assist each other in managing behavior, or there is no student misbehavior.</td>
</tr>
</tbody>
</table>
### 5D+™ Teacher Evaluation Rubric

#### Classroom Environment & Culture

<table>
<thead>
<tr>
<th>Ineffective</th>
<th>Partially Effective</th>
<th>Effective</th>
<th>Highly Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CEC6</strong> Classroom Culture: Student status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher does not develop appropriate and positive teacher-student relationships that attend to students' well-being. Patterns of interaction or lack of interaction promote rivalry and/or unhealthy competition among students or some students are relegated to low status positions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher demonstrates appropriate teacher-student relationships that foster students' well-being. Patterns of interaction between teacher and students may send messages that some students' contributions are more valuable than others.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher and students demonstrate appropriate teacher-student relationships that foster students' well-being and adapt to meet individual circumstances. Patterns of interaction between teacher and students and among students indicate that all are valued for their contributions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher and students demonstrate appropriate teacher-student and student-student relationships that foster students' well-being and adapt to meet individual circumstances. Patterns of interaction between teacher and students and among students indicate that all are valued for their contributions. Teacher creates opportunities for students' status to be elevated.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **CEC7** Classroom Culture: Norms for learning |
| Classroom norms are not evident and/or do not address risk taking, collaboration, respect for divergent thinking or students' culture. |
| Classroom norms are evident and encourage risk taking, collaboration, respect for divergent thinking and students' culture. Teacher and student interactions occasionally align with the norms. |
| Classroom norms are evident and encourage risk taking, collaboration, respect for divergent thinking and students' culture. Teacher and student interactions frequently align with the norms. |
| Classroom norms are evident and encourage risk taking, collaboration, respect for divergent thinking and students' culture. Teacher and students refer to the norms and or interactions consistently align with the norms. Students remind one another of the norms. |
### Professional Collaboration & Communication

<table>
<thead>
<tr>
<th>Professional Learning and Collaboration: Collaboration with peers and administrators to improve student learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
</tr>
<tr>
<td>Teacher rarely or never collaborates with peers or engages in reflective inquiry for the purpose of improving instructional practice or student learning.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Professional Learning and Collaboration: Professional and collegial relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
</tr>
<tr>
<td>Teacher rarely or never develops or sustains professional and collegial relationships for the purpose of student, staff or district growth. Teacher may subvert professional and collegial relationships.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communication and Collaboration: Parents and guardians</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
</tr>
<tr>
<td>Teacher rarely or never communicates in any manner with parents and guardians about student progress.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communication and Collaboration: Communication within the school community about student progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ineffective</strong></td>
</tr>
<tr>
<td>Teacher maintains minimal student records. Teacher rarely communicates student progress information to relevant individuals within the school community.</td>
</tr>
</tbody>
</table>
VITA
Thomas J. Dodd

EDUCATION & CERTIFICATIONS
The Pennsylvania State University, University Park, PA
   Doctor of Philosophy, Educational Leadership, Advisor: Dr. Nona Prestine, August 2015
   Pennsylvania Letter of Eligibility (Superintendent’s Certificate), July 2005
   Pennsylvania Administrative I Certificate (Principal’s Certificate K-12), July 2005

Colorado State University, Fort Collins, CO
   Colorado Professional Principal License K-12, June 1999

Adams State University, Alamosa, CO
   Master of Arts, Secondary Education, Social Studies emphasis, July 1995
   Master of Arts, Health, Physical Education, and Recreation, July 1995
   Colorado Professional Teacher License Social Studies 7-12, July 1995

Susquehanna University, Selinsgrove, PA
   Bachelor of Arts, History, Secondary Education minor, May 1992
   Pennsylvania Instructional I Certificate Social Studies 7-12, May 1993

EMPLOYMENT
Lesher Junior High/Middle School- an IB World School, Fort Collins, CO
   Principal, July 2005 - present
The Pennsylvania State University, University Park, PA
   Graduate Research Asst./Asst. Editor American Journal of Ed., August 2004 - June 2005
   Graduate Research Asst./Ed. Leadership Certification Officer, August 2003 - July 2004
Aspen High School, Aspen, CO
   Vice Principal, July 2001 - July 2003
Eagle Valley High School, Gypsum, CO
   Associate Principal/Head Wrestling Coach, July 2000 - June 2001
   Associate Principal/Social Studies Teacher/Head Wrestling Coach, July 1998 - June 2000
Adams State University, Alamosa, CO
   Instructor of EPLS/Graduate Assistant Wrestling Coach, August 1993 - July 1995

HONORS & AWARDS
CASSP Board Member (Region 2a Director) 2015 - 2018
CASSP/NASSP Colorado Middle Level Principal of the Year finalist 2011 & 2015
CAMLE/Natl. Forum to Accelerate Middle Grades Reform School to Watch 2014
USDE Green Ribbon School 2014
MetLife-NASSP Breakthrough School 2012
Dean’s Research Assistantship Penn State University 2003 - 2005
Pennsylvania Ed. Leadership Foundation Scholarship (PAESSP) 2004
Graham Endowed Fellowship Penn State University 2004
CASSP/NASSP Colorado Asst. Principal of the Year finalist 2003