1-1-2014

Teachers' Motivational Responses to New Teacher Performance Management Systems: An Evaluation of the Pilot of Aldine ISD's inVEST System

Claire Robertson-Kraft
University of Pennsylvania, claire.rk@gmail.com

Teachers' Motivational Responses to New Teacher Performance Management Systems: An Evaluation of the Pilot of Aldine ISD's inVEST System

Abstract
Research has shown that some teachers are dramatically more effective than others and further, that these differences are among the most important schooling factors affecting student learning. Accordingly, shifts in policy have resulted in the development of new performance management systems with the goal of improving teacher effectiveness. Although a growing body of research has begun to examine the impact of recent systems, we have very limited knowledge on how these systems influence teachers' motivation and improvement. This dissertation moves the body of research forward by using expectancy-value theory and mixed-methods analysis to examine the impact of INVEST, a new teacher evaluation system in Aldine ISD in Houston, Texas, on teacher motivation, effectiveness, and retention. It also explores how individual personality characteristics, school organizational factors, and evaluation system features influence these outcomes.

It employs a mixed methods design, utilizing the strengths of both methodological approaches. The quantitative research captures broad-based results from a teacher survey given to the population of teachers pre- and post- pilot and uses difference-in-differences analysis to examine the impact of the pilot on key outcomes (i.e., motivation, effectiveness, and retention) and multiple regression analysis to examine which predictors (at the individual, school, and system level) influenced outcomes. This analysis is supplemented by the qualitative research which draws from a small purposive sample of teachers to gain an in- depth understanding of how the policy influenced teachers' experiences.

Analyses revealed that overall INVEST had a negative impact on teachers' belief in their abilities (expectancy) and no significant impact on the importance they placed on their work (value), their effectiveness, or their decision to remain in teaching. However, teachers' responses varied considerably based on their individual characteristics (e.g., teachers' grit), their school's conditions (e.g., leadership), and their system perceptions (e.g., understanding, accuracy of measures, quality of feedback). The extensive data collected in this analysis offer a rich picture of the implementation of new performance management systems. Thus, it provides both policymakers and researchers with a better understanding of how new policies impact teacher's behavior and the influence of various characteristics (at the individual, school, and system level).

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Education

TEACHERS' MOTIVATIONAL RESPONSES TO NEW TEACHER PERFORMANCE MANAGEMENT SYSTEMS: AN EVALUATION OF THE PILOT OF ALDINE ISD'S INVEST SYSTEM

Claire Robertson-Kraft

A DISSERTATION

in

Education

Presented to the Faculties of the University of Pennsylvania

in

Partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

2014

Supervisor of Dissertation:

_____

Richard M. Ingersoll, Board of Overseers Professor of Education and Sociology

Graduate Group Chairperson:

_____

Stanton E.F. Wortham, Judy & Howard Berkowitz Professor of Education

Dissertation Committee:

Richard M. Ingersoll, Professor of Education and Sociology

Angela Duckworth, Associate Professor of Psychology

Matthew Steinberg, Assistant Professor of Education

ACKNOWLEDGEMENTS

This dissertation is the culmination of the first decade of my professional career, so there are many people I need to thank for supporting me throughout the process. I am privileged to have had an incredibly encouraging dissertation committee and a positive experience over my past five years at Penn GSE. My chair and advisor, Richard Ingersoll, always challenged me to carefully examine both sides of any issue and ensure my research was rooted in theory while simultaneously being accessible to a broader audience. Fittingly, Angela Duckworth helped me build the "grit" I needed to master the publication process and her ability to make research compelling to practitioners has been a continual source of inspiration for me over the years. As someone whose expertise entering graduate school was more aligned with qualitative research, I was fortunate to have Matthew Steinberg help me develop quantitative analysis skills so that I could become a true mixed methods researcher.

During my time at GSE, I have had many instructive learning experiences, but none was more instrumental than my Institute for Education Sciences (IES) fellowship. In addition to exposing me to new research methods and leading researchers, IES (and in particular, Rebecca Maynard) offered me the flexibility to work on projects that aligned with my interests. I had the opportunity to work on projects that were focused on Philadelphia education (with Research for Action) and rooted in psychology (with the Duckworth Lab). Indeed, my graduate experiences have helped solidify my passion for working in Philadelphia and built my understanding of the role psychology plays in influencing educational outcomes.

Prior to enrolling at GSE, I began my work in education policy at Penn with Ted Hershberg at Operation Public Education (OPE) in 2007. Over the past seven years, I edited a volume on teacher policy reform, supported a school district to design and implement a comprehensive new initiative, and evaluated the policy through this dissertation research. It is truly a unique experience to have seen the process of reform through from ideation to implementation, and throughout it all, Ted has been my biggest cheerleader, for which I am ever grateful. Though OPE has had many staff members over the years, Katie Schlesinger and Jess Yee deserve special recognition. They helped do the grunt work of research (e.g., printing, assembling, and entering data for thousands of teacher surveys) and never once complained. Everyone should be as lucky to have such conscientious co-workers.

I am also extremely grateful to my family and friends for their patience and support throughout the process. My mother, Lois Kraft, served as my own personal research librarian and read numerous drafts; my father, Alan Robertson, taught me how to use Microsoft Access (several times); my boyfriend, Paul Hughes, engaged in countless conversations on motivational theory and teacher evaluation reform; and my fellow GSE students – in particular, Jess Beaver, Nina Hoe, and Jamey Rorison who graduated alongside me – provided the much needed camaraderie to stay motivated throughout the process.

Finally, I'd like to acknowledge the Aldine ISD Leadership Team for their partnership and commitment to learning from research over the past few years and the Aldine ISD teachers who made this work possible by sharing their experiences with me.

It was during my time as a third grade teacher in Houston that I built the passion I have for working in education policy today, and it is my hope that this research will enable policymakers to design and implement teacher performance management systems that help teachers maximize student learning.

The research reported here was supported by the Laura and John Arnold Foundation and the Institute of Education Science at the U.S. Department of Education, through Grant #R305A080280 and Grant #R305B090015 to the University of Pennsylvania. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

ABSTRACT

TEACHERS' MOTIVATIONAL RESPONSES TO NEW

TEACHER PERFORMANCE MANAGEMENT SYSTEMS:

AN EVALUATION OF THE PILOT OF ALDINE ISD'S INVEST SYSTEM

Claire Robertson-Kraft

Richard M. Ingersoll

Research has shown that some teachers are dramatically more effective than others and further, that these differences are among the most important schooling factors affecting student learning. Accordingly, shifts in policy have resulted in the development of new performance management systems with the goal of improving teacher effectiveness. Although a growing body of research has begun to examine the impact of recent systems, we have very limited knowledge on how these systems influence teachers' motivation and improvement. This dissertation moves the body of research forward by using expectancy-value theory and mixed-methods analysis to examine the impact of INVEST, a new teacher evaluation system in Aldine ISD in Houston, Texas, on teacher motivation, effectiveness, and retention. It also explores how individual personality characteristics, school organizational factors, and evaluation system features influence these outcomes.

It employs a mixed methods design, utilizing the strengths of both methodological approaches. The quantitative research captures broad-based results from a teacher survey given to the population of teachers pre- and post- pilot and uses difference-in-differences analysis to examine the impact of the pilot on key outcomes (i.e., motivation,

effectiveness, and retention) and multiple regression analysis to examine which predictors (at the individual, school, and system level) influenced outcomes. This analysis is supplemented by the qualitative research which draws from a small purposive sample of teachers to gain an in- depth understanding of how the policy influenced teachers' experiences.

Analyses revealed that overall INVEST had a negative impact on teachers' belief in their abilities (expectancy) and no significant impact on the importance they placed on their work (value), their effectiveness, or their decision to remain in teaching. However, teachers' responses varied considerably based on their individual characteristics (e.g., teachers' grit), their school's conditions (e.g., leadership), and their system perceptions (e.g., understanding, accuracy of measures, quality of feedback). The extensive data collected in this analysis offer a rich picture of the implementation of new performance management systems. Thus, it provides both policymakers and researchers with a better understanding of how new policies impact teacher's behavior and the influence of various characteristics (at the individual, school, and system level).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: REVIEW OF THE LITERATURE

Introduction

Research has demonstrated that some teachers are dramatically more effective than others, and further, that these differences are among the most important schooling factors affecting student learning (Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004; Sanders & Rivers, 1996). Despite this variation in teacher effectiveness, performance management systems have historically demonstrated little or no connection between teacher evaluation results and student learning gains (Peterson, 2000; Weisberg, Sexton, Mulhern, & Keeling, 2009). Rather than rewarding excellence based on performance, two factors currently drive teacher pay raises in the vast majority of U.S. districts: years of experience and the acquisition of education credentials (Podgursky & Springer, 2006). While proponents of the single salary schedule contend that this continues to promote equity, reformers argue that teachers should not be paid based on these factors, given what we now know about the significant variability in teacher effectiveness (Hanushek, Kain, O'Brien, & Rivkin, 2005; Odden, 2008).

The U.S. Department of Education's guidelines for awarding grants from the Race to the Top Fund directly challenged the current system. To make their applications competitive, states were required to develop systems for using student growth data – as one of multiple measures – to evaluate and reward highly effective teachers. These shifts in policy have resulted in a flurry of activity surrounding the development of new teacher performance management systems. In the past few years alone, over 40 states and dozens of districts have made changes to their policies, increasing the emphasis on student

growth in teacher evaluation and ramping up the consequences attached to that evaluation. Forty-four states now require teacher ratings to be based on multiple measures of performance and 41 of these states mandate that student growth be a part of teacher evaluation systems. An increasing number of states and districts are also linking teacher evaluation results with tenure decisions and compensation reform (Doherty & Jacobs, 2013).

Unlike historical studies, recent research has demonstrated a positive, though relatively small, correlation between principal observation of teachers and student progress (Kane & Staiger, 2012; Sartain, Stoelinga, & Brown, 2011). However, the results of these new performance management systems' impact on student achievement have varied depending on how systems are designed. Studies of performance-based pay initiatives have demonstrated that bonus systems (where teachers receive a reward for students' growth) have limited to no effects on student learning (Glazerman & Seifullah, 2010; Springer et al., 2010). Conversely, several recent studies focused on more comprehensive new teacher evaluation systems demonstrate a positive impact in the early stages of implementation (Dee & Wyckoff, 2013; Steinberg & Sartain, forthcoming 2014; Taylor & Tyler, 2011).

What is unclear is why certain changes may or may not be occurring, as most of these studies do not systematically explore how teacher motivation and behavior resulted in observed outcomes. Prior research on teachers' attitudes demonstrates that their support for these types of reforms varies considerably depending on how the system is designed and implemented (Ballou & Podursky, 1993; Farkas, Johnson, & Duffett, 2003;

Goldhaber, 2009; Kelley, Heneman, & Milanowski, 2000). Though there is some research on motivational responses to accountability policies (Finnigan & Gross, 2007; Kelley et al., 2000), most studies of performance management systems do not take into consideration how design features, as well as individual and organizational characteristics, affect teacher attitudes and subsequently influence motivation.

This dissertation will move the body of research on performance management policies forward by examining the impact of INVEST, a new teacher evaluation system in the Aldine Independent School District (ISD), Houston, Texas, on teacher motivation, effectiveness, and retention, and exploring how individual personality characteristics, school organizational factors, and evaluation system features influence these outcomes. In particular, I will explore several research questions. The first research question examines the implementation of the new evaluation system and teachers' attitudes towards the policy. The second research question explores the new system's impact on teacher motivation, effectiveness, and retention. The final set of research questions investigates the relationship among all three of these outcomes (teacher motivation, effectiveness, and retention) and measures of individual personality characteristics (i.e., the Big Five, grit), school organizational factors (i.e., school climate indicators), and evaluation system features (e.g., perceptions of the measures and process). The dissertation is divided into the following chapters:

- ***Chapter 1. Review of the Literature.*** In this chapter, I provide a brief overview of the history of performance management systems, examine the empirical evidence on these systems' potential for increasing teacher quality, and finally, explore

what we can learn from theory about teachers' likely motivational responses. I develop a conceptual framework, derived from the literature on motivational theory, to frame how we might expect teachers to respond to new performance management initiatives.

- ***Chapter 2. Methods and Data Collection.*** I then turn my attention to the particulars of my proposed dissertation study and outline the three research questions I will address through my analysis. These questions fill existing gaps in the literature, particularly with regard to the impact of new evaluation systems on teacher motivation.
- ***Part One Findings: Overall***
    - ***Chapter 3. Research Question 1: System Implementation Descriptive Analysis.*** In this chapter, I share descriptive data on system implementation and explore trends in teacher attitudes. I then provide an overview of variation at the individual and school level.
    - ***Chapter 4. Research Question 2: Overall System Impact.*** After presenting the descriptive results, I evaluate the impact of the new INVEST system on teacher motivation, effectiveness, and retention. I examine quantitative data analyzed through the difference-in-differences approach to estimate the treatment effect and supplement this quantitative analysis with qualitative data gathered through teacher interviews.
    - ***Part Two Findings: VariationChapters 5, 6, and 7. Research Question 3: Variation in Implementation and Impact.*** In these chapters, I explore

how variation in individual characteristics (Chapter 5), school characteristics (Chapter 6), and system characteristics (Chapter 7) influence the outcomes discussed in Chapter 4. I use multiple regression analyses to examine which factors best predict outcomes of interest – e.g., teacher motivation, effectiveness, and retention – and use the qualitative data to explain these trends.

- ***Chapter 8. Discussion and Implications.*** To close, I revisit the framework developed in Chapter 1 for understanding the impact of new systems on teacher motivation, effectiveness, and retention. With this framework in mind, I discuss the various implications of my work for policymakers and practitioners and identify areas for further research.

Research Overview

The Need for New Performance Management Systems

Broadly speaking, performance management systems aim to address the problem of teacher quality. Over the past decade, a growing body of research evidence has demonstrated that teacher effects can have a substantial impact on student progress (Chetty, Friedman, & Rockoff, 2011; Gordon, Kane, & Staiger, 2006; Kyriakides & Creemers, 2008; Rockoff, 2004). Unfortunately, teachers vary considerably in their effectiveness, and students from low-income families are less likely to have access to high quality instruction than their peers in higher-income communities (Walsh, 2007). The problem of teacher quality is multi-faceted and, consequently, policymakers have come to understand it in different ways. Some argue that policy should focus on

attracting more high quality candidates into the profession and encouraging them to teach in schools with the highest need. Others contend that policymakers conceptualizing the problem solely as one of recruitment fail to recognize that the shortage is not a result of too few quality teachers entering the profession, but rather is exacerbated by the alarming proportions in which they leave. And yet others assert that if the system cannot accelerate teachers' improvement or maximize their potential, recruiting and retaining more teachers will not adequately address the issue. Thus, the "problem" of teacher quality can be conceptualized as one of inadequate recruitment, high turnover, or a lack of improvement (Johnson, Berg, & Donaldson, 2005).

Historically, teacher performance management systems have not been intentionally designed to respond to any of the conceptions of the teacher quality problem and thus do not meaningfully differentiate performance or reward excellence (Peterson, 2000). Indeed, in *The Widget Effect*, The New Teacher Project researchers discovered that more than 99% of teachers in examined districts were rated satisfactory and that this tendency had fostered an environment where policymakers treat teachers as interchangeable parts (Weisberg et al., 2009). To respond to these shortcomings, reforming teacher performance management systems (i.e., evaluation, compensation, support, dismissal) has become central to policy conversations at the national, state, and local level.

Advocates of these new systems argue that better differentiating performance and aligning consequences directly with outcomes will address the "teacher quality problem" through both a selection and a motivation effect. A system which aligns performance and

rewards will attract individuals who are particularly skillful at the outcome being rewarded, and this selection effect will have a positive impact on the labor market (Podgursky & Springer, 2006). Clear performance expectations and aligned incentives will in turn motivate current teachers to change their behaviors and remain in the profession (Odden & Wallace, 2007). For the purposes of this analysis, I will focus specifically on the motivation effect of new performance management systems on the existing teacher corps. This is not to suggest that the selection effect is not an equally important outcome to consider, and future work should certainly explore the effect these initiatives have on potential recruits.

Key Elements of New Performance Management Systems

Various forms of performance management have come and gone in waves over the years. In the early 1900s and then again in the 1950s and 1980s, policymakers designed new merit-based pay systems to improve teacher quality, largely in response to fear over intensified international competition. Despite their initial popularity, the evaluation criteria in these systems were perceived as subjective, and they subsequently failed to engender broad-based support. Additionally, districts faced considerable implementation challenges including difficulties in reliably training evaluators, union opposition, instability in leadership, and a lack of sustainable funding (Johnson, 1984). Largely structured as top-down initiatives, these programs neglected to secure support from influential constituencies such as teachers and without a clear rationale for why rewards were disseminated to some teachers and not others, policies engendered low morale (Cohen & Murnane, 1985; Darling-Hammond & Berry, 1988). Combined with

funding challenges and lack of sustained leadership, performance management initiatives have historically been transient in nature (Johnson, 1984).

In an era of high stakes accountability, policymakers face intensified pressure to improve test results and consequently an increasing number of districts are again in the process of developing performance management systems (Doherty & Jacobs, 2013; Podgursky & Springer, 2007). These efforts have been accelerated by the U.S. Department of Education's Race to the Top Fund guidelines released in 2009 and subsequently by the No Child Left Behind waiver requirements. To make their applications competitive, states were required to develop new systems that addressed teacher evaluation, compensation, and professional development. The fundamental aim of these new systems is to provide a mechanism for differentiating teacher effectiveness for accountability purposes, while simultaneously driving improvements in practice. To accomplish this goal, advocates have called for a balanced approach, using multiple measures to gauge teacher effectiveness and recognize outstanding performance (Aspen Institute, 2011).

Though these new systems vary considerably, most share a number of core design features. First, they use multiple measures of teacher performance – typically a student growth or value-added model and a robust observation framework. To respond to the shortcomings of previous attempts at measuring teachers' impact on students, value-added models attempt to control for the other school- and student-based factors influencing outcomes, thus isolating the impact of the teacher on student progress (Goe, 2008; Lockwood & McCaffrey, 2008; Meyer & Christian, 2008). On the observation

side, new systems employ comprehensive frameworks that capture a more complete picture of teaching behaviors than previous observation systems, differentiate performance across a number of levels, and provide timely and detailed feedback about specific teachers' strengths and areas for improvement (Milanowski, Heneman, & Kimball, 2009). Additionally, these performance management systems tend not to be focused on evaluation alone, but rather are part of a more comprehensive approach, including other reforms with the objective of increasing teacher quality (e.g., compensation, professional development) (Odden & Wallace, 2004).

Empirical Evidence: What Do We Know about These Systems' Impact?

Designing new performance management systems has been at the heart of education reform efforts for the past century; yet, surprisingly little information exists about how these new approaches work in practice. The basic logic undergirding these systems is that through improved evaluation, policymakers will be able to better identify highly effective and ineffective teachers, as well as capture important information on all teachers' areas of need. Policymakers can then use this knowledge to design specific policy interventions – e.g., pay for performance, enhanced professional development, remediation for struggling teachers, dismissal of ineffective teachers – that will build both teacher motivation and capacity and ultimately, improve the quality of instruction.

Determining Validity and Reliability of Measures

A considerable amount of the research on these new systems has focused on the validity and reliability of the performance measures. History has made clear that defining high quality teaching is an unusually challenging task because it requires making

judgments on an issue for which there is considerable disagreement. Many scholars contend that quality teaching takes on different characteristics in different contexts and as a result, good teaching does not lead to successful teaching absent the right conditions for learning (e.g., student engagement, parental support, sufficient resources) (Berliner, 1976; Fenstermacher & Richardson, 2005). Thus, developing measures of performance is particularly challenging in education because goals are complex and effective instruction cannot be attributed to the teacher alone (Harris, 2011; Kelly, 2011). In an attempt to address this concern, most new performance management systems employ multiple measures. Below, I will draw from the empirical literature to investigate the validity and reliability of these various measures for use in high-stakes contexts.

Value-Added. Proponents of value-added models (VAMs) contend that these modeling techniques control for other factors influencing outcomes, and thus can isolate the impact of the teacher on student learning (Goe, 2008; Meyer & Christian, 2008). Though the use of VAMs continues to receive attention, research on the validity of these measures is quite polarized. Some researchers caution that measuring teacher effectiveness through student test score gains has significant methodological and practical challenges (Baker et al., 2010; Rothstein, 2008; Rothstein, 2009), while others contend that despite limitations, these measures are the best predictors we have about future student performance (Glazerman et al., 2010; Kane & Staiger, 2012). These debates center around the value we should place on students' test scores as a measure of performance and the extent to which student growth offers a valid and stable measure of teacher effectiveness.

The first set of researchers' concerns deals with how best to assess student performance. At the most basic level, different tests measure different content, and some researchers have questioned whether existing assessments truly measure outcomes we value. In a recent study, Jennings and Corcoran found that the teacher effect is 15-30% larger on the high stakes test than on low stakes tests, suggesting that teacher effects may not persist across assessments (Jennings & Corcoran, 2011). In another analysis, they discovered that while teacher effects on math and reading value-added scores were highly correlated, correlations with social/behavioral skills tended to be much lower, implying that value-added outcomes may not be strongly associated with other measures believed to lead to long-term success (Jennings & Corcoran, 2011). Conversely, a recent analysis discovered that students assigned to higher value-added teachers were more successful over the long-term and had higher rates of college attendance, more substantial salaries, and better life outcomes (Chetty et al., 2011).

Regardless of whether test score growth predicts other valued outcomes, researchers have also raised concerns over the validity and reliability of value-added measures when used for high stakes purposes. Most notably, students are not randomly distributed across classrooms, and selection into classrooms based on unobservable characteristics (e.g., principals' sorting of teachers based on unobserved student characteristics) could bias results (Rothstein, 2008; Rothstein, 2009). Though this is an inherent limitation of value-added measures, several studies have suggested that the selection based on unobservables is small and that the quality of teaching (as measured

by value-added assessment) does not differ systematically across types of schools and students (Kane & Staiger, 2008).

Researchers have also raised questions about the extent to which value-added estimates can provide a reliable inference about a teacher's effectiveness (Koedel & Betts, 2007). Several studies have demonstrated that value-added estimates for teacher-level analyses are subject to random error (Lockwood & McCaffrey, 2008; Schochet & Chiang, 2010). Others recognize these limitations but contend that the stability of VAMs is comparable to standards of evaluation in other fields and provides a more reliable picture of teacher performance than existing indicators (Glazerman et al., 2010; Kane & Staiger, 2012). As recent research has made clear, the specifics of how growth models are constructed (e.g., whether they control for individual and/or school covariates) can yield different results on both teacher and school effectiveness (Ehlert, Koedel, Parsons, & Podgursky, 2013).

Teacher Observation. Skeptics of using value-added assessment believe teaching is more complex than can be captured by student performance on standardized assessments and argue that teachers should be assessed based on their actions, not just their outcomes. In response, many states and districts are now employing more sophisticated teacher performance assessment systems as the basis of high-stakes decisions (Milanowski et al., 2009; Gallagher, 2004; Jacob & Lefgren, 2008). Recently, researchers at the Gates Foundation reviewed several such systems through the Measures of Effective Teaching Project – e.g., The Framework for Teaching developed by Charlotte Danielson – and discovered a positive, though relatively small, correlation

between observation results (conducted by external raters rather than principals) and student learning (Kane & Staiger, 2012).

When used in high-stakes environments, researchers have contended that observation measures should be viewed as systems, not merely instruments (Hill, Charalambous, & Kraft, 2012). To maximize reliability, evaluators should receive adequate training in the evaluation system and demonstrate their competency level before decisions are used for high-stakes outcomes (Hill et al., 2012; Kane & Staiger, 2012). However, inter-rater agreement, while important, should not be the sole reliability metric. Indeed, teaching behavior can vary from day to day and week to week, meaning that one observation is unlikely to provide an accurate view of teacher performance, particularly if it is announced and the teacher can prepare in advance. Recent research has demonstrated that reliability can only be achieved through multiple observations of practice (Hill et al., 2012; Kane & Staiger, 2012), and unfortunately, some evidence suggests that using principals as the primary evaluators can lead to leniency and limit score differentiation (Milanowski et al., 2009; Weisberg et al., 2009). In short, though there has been considerable research focused on these performance measures, much remains to be learned about their validity and reliability. Although these new measures may be able to better differentiate between teachers' practice, researchers should continue to closely monitor how they impact teachers' motivation and in turn influence their effectiveness.

Impact

Teacher evaluation tools should not only be assessed on their ability to accurately differentiate teacher performance, but also on how well they inform and support teacher

development. As discussed above, much of the current research on performance management systems has focused on the validity and reliability of various measures, yet considerably fewer studies have examined the impact these systems have on teacher effectiveness and, in turn, student progress. To complicate matters, the growing body of rigorous research that does exist reveals mixed results. This section will examine the existing literature and explore possible explanations for the discrepancy in findings across studies.

In their 2006 review, Podgursky and Springer reported on rigorously conducted studies employing a treatment and control design and found that in most instances, performance incentives were associated with increased student achievement. Because treatments varied considerably from study to study, conducting a meta-analysis was not possible, but the majority of studies examined found that the incentives had a direct effect on the variable being incentivized. Specifically, Lavy (2007) investigated a tournament designed to raise pass rates on high school exit exams in low socioeconomic status high schools in Israel. Teachers participating in the program were ranked based on exit exams and received substantial bonuses. At the close of the year, participant teachers' performance increased when compared to control teachers. In their study of the impact of similar systems in the United States, Figlio and Kenny (2007) analyzed data from the national cross-sectional analysis on schools, students, and families and discovered that test scores were higher in schools that offered individual financial incentives for good performance.

Several other evaluations have discovered positive outcomes. A study by Dee and Keys (2004) examined the relationship between teachers' evaluation results (and corresponding placement on a career ladder) and student achievement gains using Tennessee Project STAR data. They found that teachers with higher status were more effective, as measured by gains in student progress. In Little Rock, researchers used a difference-in-differences approach to analyze the impact of a new performance management system and discovered that students of participating teachers made larger test score gains than students taught by teachers in the comparison group (Winters, Greene, Ritter, & Marsh, 2008). A similarly positive effect was found among teachers who opted to participate in the Denver ProComp program, which differentiated teacher compensation based on a variety of performance measures (Wiley, Gaertner, Spindler, & Subert, n.d.).

However, other research on performance incentives has suggested the opposite to be case. In the first randomized control study of performance pay initiatives ever conducted in the United States (of the Project on Incentives in Teaching – POINT – experiment in Nashville), researchers found that teacher performance pay did not raise student test scores. Teachers were eligible for up to $15,000 as an incentive and lesser amounts were rewarded for lower thresholds. The only effect was observed in fifth graders taught by teachers who received bonuses, but the gains in student achievement did not persist into the subsequent year (Springer et al., 2010). Another recent evaluation study conducted on the Teacher Advancement Program, where schools were randomly assigned once they had volunteered to participate in the program, also discovered no

evidence that the performance management system increased student achievement or teacher retention (Glazerman & Seifullah, 2010).

More recently, studies of new teacher evaluation systems in Cincinnati, Washington, D.C., and Chicago have yielded positive outcomes even in the early years of implementation. In Cincinnati, Taylor and Tyler (2011) found that students taught by teachers after they participated in the pilot of the Danielson Framework for Teaching scored about 10% of a standard deviation higher on standardized math achievement tests than similar students in the pilot period. Dee and Wyckoff (2013) employed a regression discontinuity design to evaluate the effect of Washington, DC's IMPACT system on low-performing teachers whose ratings placed them at the threshold (that would result in dismissal) and high-performing teachers (whose ratings meant they received a large financial incentive). Results indicated that dismissal threats increased the voluntary attrition of low-performing teachers by 11 percentage points and improved the remaining teachers' performance by .27 of a teacher-level standard deviation. Higher performing teachers at the threshold were also considerably more likely to improve their performance. In a randomized control study of Chicago Public Schools' Excellence in Teaching Project, Steinberg & Sartain, forthcoming 2014) discovered that schools piloting the new evaluation system performed better in reading and math than non-pilot schools during the pilot and subsequent year. These effects were particularly salient in higher achieving and lower poverty schools.

Why the discrepant results? For one thing, the direct evaluation literature on performance management systems is highly diverse in terms of methodological rigor.

Some studies are purely observational and do not attempt to control for other confounding variables that may impact results. To complicate matters, participation in many programs is voluntary, which means any observed effect could be due to the characteristics of those teachers who opt into the program. In these cases, it is not possible to separate the selection effect of those choosing to participate from the impact of the program itself (Podgursky & Springer, 2007).

But perhaps more importantly, the system's role in improving performance is complicated by the fact that initiatives vary considerably in their design (Johnson & Papay, 2009). Some of the initiatives discussed above are solely performance pay systems, which are fundamentally different in their design compared to more comprehensive systems rooted in improving teacher practice. Taylor and Tyler (2011) distinguish between investment in human capital and short-term accountability effects as two possible goals of policies. They contend that the effects of a system will be more likely to persist if the evaluation spurs employees' investment in human capital. The early findings from Washington, DC's IMPACT evaluation suggest that reforms with significant consequences both in the positive direction (additional pay) and in the negative direction (threat of dismissal) can also impact teacher behavior. Given the many ways programs could be designed, simply knowing whether new performance management systems have an impact on teacher and student outcomes does not provide the information necessary to understand the nature of this impact. Despite decades of interest, there is only limited research on teachers' perceptions of different system design features and why different system designs yield differing results.

To truly understand the impact of new performance management systems, researchers must also investigate how teachers' responses to new policies are influenced by individual characteristics and school organizational factors. Existing studies have demonstrated that teachers' responses to new systems vary considerably (Goldhaber, DeArmond, & DeBurgomaster, 2007), yet there is limited systematic research on teacher motivation in response to new policies. Research needs to move beyond exploring how the general pool of teachers feels about new systems to begin to understand how new evaluation systems affect teacher motivation and how this motivation varies across subgroups of teachers working in different types of contexts.

Conceptual Framework: Understanding Teacher Motivational Responses

In this analysis, I draw from a substantial body of motivational literature to develop a conceptual framework for better understanding the factors influencing teacher responses to performance-management policies and how these responses translate into instructional improvements. Originating with Vroom (1964), expectancy-value theory posits that individual performance in an organization is a function of ability and motivation (Lawler, 1983; Vroom, 1964). Motivation, or the process governing the choices individuals make, is influenced by the value of certain outcomes and the perceived relationship between actions and outcomes. In other words, how individuals initially respond to performance management policies can best be understood in terms of two sources of motivation – the desirability of a particular outcome and a person's belief that with increased effort, they can achieve that outcome (Vroom, 1964).

As discussed by Achtziger and Gollwitzer (2010), initial motivation is distinct from the volition required to sustain changes in practice. To achieve goals, individuals must shift from a deliberative to an implemental mindset and engage in self-regulatory planning. Ultimately, achieving expertise is the end result of individuals' prolonged efforts to improve performance while negotiating motivational and external constraints (Ericsson, Krampe, & Tesch-Römer, 1993). In short, individuals need to be motivated to change behavior, design initial plans of action, and then consistently and strategically work to improve performance.

Initial Motivation: An Overview of Expectancy-Value Theory

Eccles, Wigfield, & Schiefele (1998) have elaborated the general expectancy-value theory into a more comprehensive theoretical model linking motivational choices to two sets of beliefs: the expectation of success that an individual has and the importance or value the individual associates with various activities. At its most basic level, this expectancy-value model can be reduced to two central questions: "Can I do the task?" and "Do I want to do the task?" Though the focus of this work has been on students, the same general principles can be applied to teachers. If teachers do not think they are capable of achieving the expectations, they will be unlikely to change their behavior. Further, teachers who believe they can make necessary changes but do not value the task itself or the outcomes associated with the task are also unlikely to alter their motivational responses.

Expectancy. Historically, expectancy perceptions are said to be governed by the expectation that a given performance will produce particular outcomes (Vroom, 1964).

Bandura, who has written extensively on individual motivation and behavioral change, emphasizes the importance of distinguishing between these more traditional outcome expectations and perceived self-efficacy. General expectancies about the effectiveness of effort (i.e., outcome expectations) document whether an individual thinks a given behavior will lead to certain outcomes. To the contrary, self-efficacy captures a person's belief about his/her own level of competence in a particular situation. Though both are important to consider, Bandura's research demonstrates that self-efficacy better predicts performance outcomes (Bandura & Locke, 2003). Believing that actions can result in outcomes will not necessarily lead an individual to sustain personal effort in the face of a specific challenge (Bandura, 1977). More recent research on expectancy value models (Eccles et al., 1998) has similarly focused on self-efficacy perceptions (i.e., "Can I do the task?") and discovered they lead to improved student performance and motivation to take on more challenging tasks. Research has also demonstrated that self-efficacy consistently predicts levels of student achievement. In other words, more efficacious teachers produce stronger gains in student achievement than teachers with lower efficacy (Goddard, Hoy, & Woolfolk Hoy, 2004; Tschannen-Moran,Woolfolk Hoy, & Hoy, 1998).

To build self-efficacy, individuals need to receive consistent information about how their performance relates to a specific set of standards (Bandura, 1982; Bandura & Schunk, 1981). This form of proximal goal-setting provides individuals with immediate feedback on their performance related to expectations (Bandura & Locke, 2003). Achieving these interim goals leads to increased satisfaction, which, in turn builds interest in the task itself (Bandura & Schunk, 1981). Some evidence suggests that

feedback framed as gains towards goals can better sustain motivation than negative feedback, which has the potential to reduce individuals' level of expectancy. However, researchers have also determined that individuals react differently to negative feedback depending on prior levels of self-efficacy (Gist, 1987). In other words, individuals higher in self-efficacy will be more likely to set ambitious goals and respond positively to negative feedback by attributing failure to actions within their control and focusing efforts on improving performance (Bandura, 1993; Bandura & Locke, 2003).

Value. To be motivating, individuals must not only believe they can make changes in their behavior but also value the process and/or outcomes associated with increased effort. Eccles and colleagues contend that the perceived value of any given activity can be determined by four constructs: (1) the intrinsic interest one expects to get from a specific task; (2) attainment value, or the extent to which a task is consistent with an individual's self-image; (3) the utility value of the task for achieving long-range goals, and (4) the perceived cost of a particular action (Eccles, 2007; Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006). These same constructs provide a useful framework for considering the value teachers place on new performance management systems.

Intrinsic value refers to the interest an individual takes in executing a given task. Individuals' intrinsic interest is maximized when they are pursuing tasks that are enjoyable and aligned with their personal preferences. While everyone may agree that certain tasks are inherently interesting, some individuals will inevitably be more likely to find specific tasks (e.g., sports, arts) more interesting than others. Psychologists have also

demonstrated that regardless of the specific task, individuals are intrinsically motivated to fulfill basic human needs (Wigfield et al., 2006). In particular, self-determination theorists have demonstrated that activating the basic psychological needs of autonomy (our desire to be causal agents of our own lives), competence (our desire to experience mastery), and relatedness (our desire to interact and be connected to others) fosters higher levels of value for particular tasks. In the case of performance management systems, some teachers may receive inherent enjoyment from being competent or feeling valued by others, which will motivate them to work harder to meet performance targets. However, Deci and Ryan (2000) would argue that this intrinsic value is only activated if teachers feel they have control over their own actions under new systems.

Even if individuals are not intrinsically interested in specific tasks, they can still find value in their long-term benefits – i.e., attainment or utility value (more generally understood as extrinsic motivation). Attainment value is the link between specific tasks and individuals' needs and identities, while utility value refers to whether the task will help individuals achieve their long-term goals (Eccles et al., 1998; Wigfield et al., 2006). In the case of performance management policies, teachers who want to be perceived as effective in their role by others will place higher value (i.e., attainment value) on reaching performance targets. Additionally, those who desire to move into a leadership position within their school will likely be more motivated to achieve greater recognition (i.e., utility value).

When determining whether to act, motivational theorists contend that individuals will weigh the value (i.e., intrinsic interest, attainment, and utility value) with perceived

costs. Cost can be affected by any number of factors, including anxiety about failure or the perceived loss of time for activities that are of greater interest (Eccles et al., 1998; Wigfield et al., 2006). In the context of new performance management systems, teachers might not desire recognition for fear of creating animosity among their colleagues and jeopardizing their ability to collaborate in meaningful ways. Alternatively, they might value being perceived as competent but opt instead to spend more time with their individual families for whom they have greater interest and commitment.

In sum, teachers' motivation will be a function of their expectancy and the value associated with specific performance outcomes. Teachers must believe they can achieve the expectations or task at hand and believe that doing so will result in something of value, either an immediate sense of satisfaction or a step in the right direction toward achieving a long-term benefit.

Factors Affecting Motivation

Expectancy-value theory posits that individuals' motivational responses to external influences will be a function of both personal factors and environmental conditions (Bandura, 1977). In other words, not all teachers will respond to the same policies in an identical fashion. Indeed, teachers' motivational reactions to new performance management policies are likely influenced by perceptions of the system, as well as differences in individual characteristics and school-based factors.

Perceptions of System Features. Teachers' perceptions of new systems will be influenced by their level of understanding of – and the value they place on – the principles undergirding the new system. According to expectancy-value theory, goals will

only be motivating if they align with individual values, so teachers must believe they will gain some sort of intrinsic enjoyment from achieving results or that reaching higher levels of performance will lead to longer-term benefits. To maximize motivational responses, teachers must value performance metrics and believe they are accurate perceptions of their performance. Additionally, theory makes clear that an individual's motivation is strengthened when performance goals are clearly defined. This clarity allows individuals to determine the value they attribute to particular goals and how likely they are to achieve them with increased effort (Locke & Latham, 1990). If systems become too complex, they run the risk of resulting in a lack of clarity and corresponding decrease in motivation.

Individual Characteristics. To be motivating, performance management systems must be congruent with the expectancies and preferences of the individuals they are designed to impact. Given this, we should expect motivational responses to performance management policies to vary across subgroups of teachers – in particular, by years of experience, effectiveness, and personality. Researchers have demonstrated that self-efficacy increases with demonstrated success (Bandura, 1977; Gist, 1987) and further, teachers improve their effectiveness considerably in the first few years in the profession (Hanushek, 1996; Hanushek & Rivkin, 2004). As a result, many novices will likely have lower levels of expectancy than more experienced teachers. Similarly, since highly effective teachers will have achieved greater success in the classroom, they are also likely to have higher expectancies regarding their abilities to meet new performance outcomes. Research has also demonstrated that individual differences in teacher personality

influence teachers' level of engagement in their work (Teven, 2007) and attitudes towards the implementation of new systems (Somech, 2010). Although many personality inventories exist, the five-factor theory – emotional stability, extraversion, conscientiousness, agreeableness, and openness to experience – has emerged as the foundational approach to describing personality traits (Goldberg, 1990**;** John & Srivastava, 1999; McCrae & Costa, 1987). It is likely that certain Big Five traits influence teachers' responses (e.g., teachers who are more open to new experience may be more receptive to change).

Organizational Factors. Research on levels of expectancy in schools has demonstrated that teachers' sense of efficacy can also be influenced by school-level variables. The most prominent of these factors include the presence of a professional community, the quality of principal leadership, and the level of teacher involvement in decision-making structures (Kelley et al., 2000; Rosenholtz, 1989). Researchers have discovered that professional community can be a strong predictor of teacher expectancy, as teacher efficacy beliefs are higher in schools where teachers work collaboratively to enhance practice (Tschannen-Moran et al., 1998). Effective principals are able to create a clear vision for success and invest teachers in a common purpose, thus deepening the sense of professional community and increasing expectancy perceptions. Rather than creating a top-down culture, effective principals offer teachers meaningful involvement in the decision-making process, which in turn, increases the value they place on policies (Ashton & Webb, 1986; Tschannen-Moran et al., 1998)**.**

Effectiveness: Translating Motivation into Improved Performance

Even if teachers are motivated to increase effort, expectancy-value theory does not posit that this alone will lead to improvements in performance. Indeed, this initial motivation must be translated into actions designed to impact practice and then these actions must be sustained over time. Goal setting (a product of initial motivation) and goal striving (resulting from volition) are governed by distinct psychological processes. As described by Achtziger and Gollwitzer (2010), when individuals move from the deliberation (or goal-setting) to the action (or goal-striving) phase, they commit to a specific goal and develop implementation intentions for translating that goal into action. A substantial body of literature has demonstrated that goals are achieved when accompanied by planning for particular action and changes to practice.

Merely practicing, however, does not lead to maximal performance. Instead, according to psychologist Anders Ericsson, who has studied the development of expertise, individuals must engage in deliberate practice to improve performance. Unlike traditional practice, deliberate practice requires working at the edge of one's abilities, receiving immediate feedback on performance, and repeatedly executing the same or similar tasks. Individuals acquire expertise gradually, and new challenges must take into account pre-existing knowledge, as well as be scaffolded and sequenced over time. Engaging in deliberate practice requires intense concentration in the face of challenge (Ericsson, 2006) and immediate and specific feedback to accelerate the growth process (Ericsson, Nandagopal, & Roring, 2009). This type of practice, though not pleasurable,

has resulted in the development of expertise across a variety of different fields (Ericsson et al., 1993).

Perceptions of System Features. Both theory and research demonstrate that the effectiveness of performance management systems will ultimately depend on how well they are implemented within a particular context. Goals will be more motivating when workers not only value the performance criteria but also receive consistent information about how their performance relates to a specific set of standards. Setting and achieving interim goals increases motivation and in turn, builds interest in the task itself (Bandura, 1982). In other words, evaluation cannot lead to improvements in performance unless teachers receive meaningful feedback and consistent support to implement necessary changes in their practice. Research has also demonstrated that individuals' motivational responses can be influenced by their level of participation in the decision-making process. Increased involvement builds trust and engenders overall commitment to new systems (Lawler, 1983).

Individual Characteristics. Research has demonstrated that certain individuals will be more predisposed to sustain motivation and, thus, improve practice over time (Achtziger & Gollwitzer, 2010). Because teaching is extremely challenging work, it seems logical that grit, defined as perseverance and passion for long-term goals, would have an important impact on teachers' volition. Two separate studies have shown that grit predicts teaching performance indexed as the academic gains of teachers' students. The first study used a self-report questionnaire (Duckworth, Quinn, & Seligman, 2009) and the second developed a résumé coding process to capture evidence of grit in college

extracurricular activities (Robertson-Kraft & Duckworth, 2014). Mediation analysis confirms that the effect of grit on outcomes is through cumulative effort: gritty individuals tend to work harder than their peers, and they remain committed to chosen pursuits over sustained periods of time (Duckworth, Kirby, Tsukayama, Berstein, & Ericsson, 2010). Gritty individuals not only show up, but they deliberately set long-term objectives and maintain effort towards achieving them, even in the absence of positive feedback (Duckworth, Peterson, Matthews, & Kelly, 2007). Following this logic, we would expect gritty teachers to remain committed to their students, set long-term objectives for the year and beyond, and sustain efforts toward improving their practice to reach these objectives.

Organizational Factors. In addition to being influenced by individual differences, teachers' ability to sustain improvements in practice is a function of their working environment. Engaging in deliberate practice is incredibly challenging and at least in the early stages, virtually impossible to do alone. Indeed, in order to successfully improve practice, teachers need consistent feedback on their performance. Given the design of new evaluation systems, the principal is most likely responsible for providing this type of support, though peer colleagues offer another possible source of coaching. According to theory, support will be most effective when it is provided on a targeted individual basis, but structures for professional learning may also have the potential to accelerate teacher improvement.

Retention: Avoiding Burnout and Staying Committed to the Profession

To sustain commitment to the profession over time, teachers must maintain initial motivation and avoid experiencing burnout. In the psychological literature, job burnout has been a critical concept for understanding individual's work experiences. Over time, individuals who experience burnout fail to sustain the hard work necessary to have a meaningful impact. In general terms, burnout is defined as "a state of exhaustion in which one is cynical about the value of one's occupation and doubtful of one's capacity to perform" (Maslach, Jackson, & Leiter, 1996, p. 20). It is characterized by emotional exhaustion, negative perceptions and feelings about clients or patients, and a crisis in professional competence (Schaufeli, Leiter, & Maslach, 2009).

Burnout is a three-dimensional construct of exhaustion, cynicism, and inefficacy, and the opposite of engagement, which includes energy, involvement, and efficacy. When energy translates into exhaustion, individuals feel fatigued when they even think about having to go to work, and the costs associated with increased job expectations do not appear worthwhile. This exhaustion stems from the fact that individuals no longer feel optimistic or involved in their work and consequently, exerting additional effort seems futile. Individuals reduce their initial expectancies when they realize they cannot make their desired impact, which in turn, can feel like an attack on their professional identity. With their sense of competence challenged, individuals decrease the value they place on their work and are generally less likely to persist over time (Maslach et al., 1996). Of course, burnout is not the only factor that influences turnover. However, it may be associated with the implementation of a new evaluation system that considerably

increases expectations for teachers and is thus a relevant construct to examine in the context of this analysis.

Perceptions of System Features. Research demonstrates that two distinct system-level factors contribute to burnout – the imbalance of demands over resources and a conflict in values between the employee and employer (Schaufeli et al., 2009). When employers place increased demands on employees without additional support, it can lead to intensified burnout, particularly when available resources are insufficient to meet the additional requirements. Employees' frustration with a potential lack of resources worsens when there is value conflict. In other words, if individuals do not share the same values as their organizations, this lack of alignment intensifies burnout experiences and leads to higher rates of employee turnover.

Individual Characteristics. Burnout is not a negative disposition, but rather the erosion of a level of positive engagement. Burnout research originated in the 1970s to examine the psyche of the idealistically motivated young people who had entered human services professions but over time became disillusioned by the systemic factors that stood in the way of their ability to make an impact. This "frustrated idealism" characterized the burnout research, as individuals lost both their energy and sense of value for their work. This experience is not unlike the plight of the urban teacher who enters the profession eager to make an impact and confronts the challenges associated with educating disadvantaged populations. Given this, we may expect to see some burnout among novices who have a particularly low threshold for challenge (i.e., low grit). Additionally, research has also discovered that individuals experience burnout when they feel the level

of recognition is not commensurate with their hard work; indeed this "lack of reciprocity," as termed by Schaufeli et al. (2009), has been shown to foster burnout. As a result, we would also expect more seasoned veteran teachers who continue to work hard year after year but feel less recognized for their efforts to experience burnout.

Organizational Factors. Teachers' long-term engagement in their work and ultimately, their decision to remain in the profession can be affected by a variety of working conditions. Many of these factors are similar to those influencing initial expectancy, including the presence of professional community, the quality of administrative support, and the level of faculty influence. Indeed, researchers have shown that increased opportunity to collaborate with colleagues can sustain teacher engagement, while principals play an essential role in maintaining teacher morale and preventing burnout in the face of significant challenge (Johnson et al., 2005). Moreover, teachers' satisfaction and subsequent decision to remain in the profession is positively associated with measures of autonomy and faculty influence (Ingersoll, 2001; Ingersoll, 2006). Additionally, teachers have cited a variety of sources contributing to their dissatisfaction, e.g., unsafe environment, inadequate resources, challenging teaching assignments, and intrusions on instructional time (Ingersoll, 2001), all of which contribute to a mismatch between demands placed on teachers and appropriate resources. Of course, gritty individuals may persist even in the fact of these challenges, but, in the aggregate, teachers' ability to sustain initial motivational responses and avoid experiencing burnout will likely be influenced by their level of satisfaction with the school environment.

Summary: Conceptual Framework Derived From Motivational Theory

In sum, expectancy-value theory provides a useful framework for examining the impact initiatives have on teachers' responses to new system. To alter teacher motivation, policies must influence teachers' expectancy that they can reach specific targets ("I can") and build the value associated with achieving certain levels of performance ("I want").

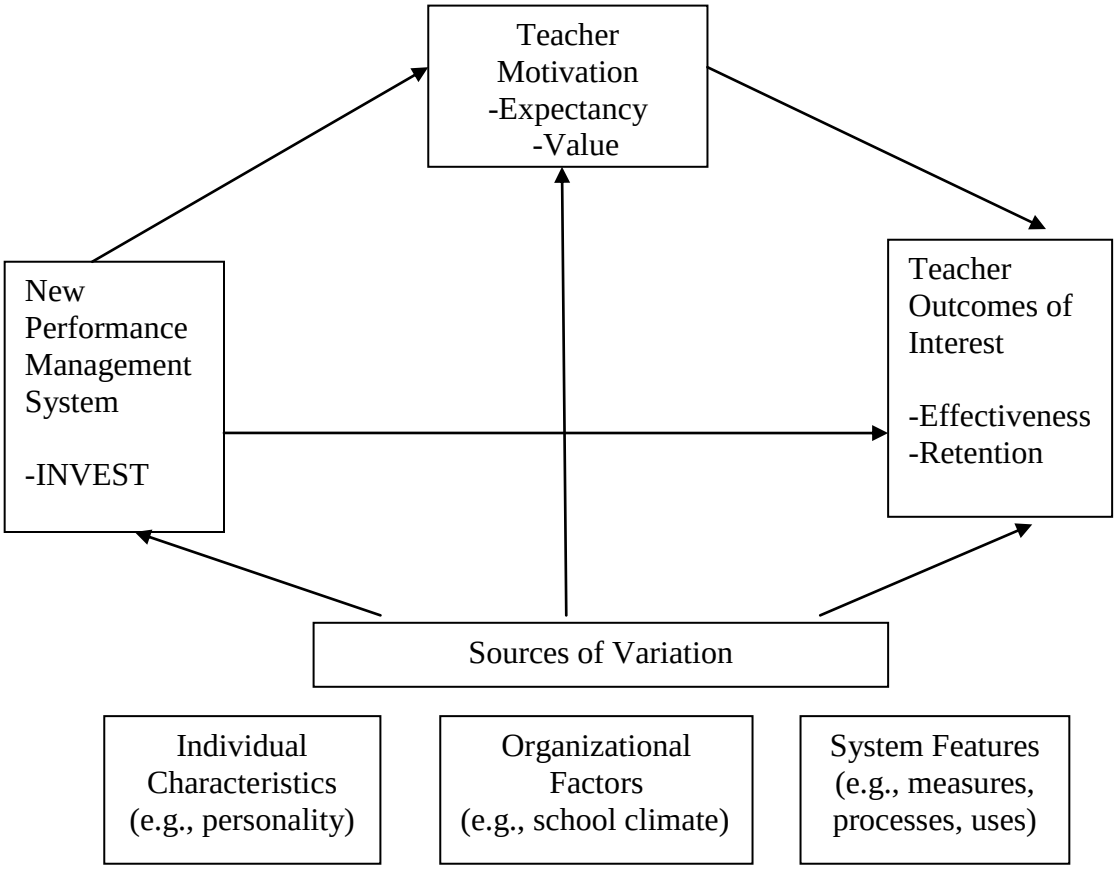

Figure 1-1. Expectancy-value framework for understanding teacher motivation

To sustain changes in practice, they must subsequently support teachers to engage in implementation planning and provide the targeted and consistent feedback necessary to improve practice and sustain commitment over time. It is essential that researchers investigate how teachers' perceptions of system features, as well as their individual characteristics and school-based organizational factors, affect both initial motivation, sustained volition, and commitment. See Figure 1-1 for an explication of how expectancy-value theory interprets teachers' reactions to new performance management policies and the impact that these reactions have on subsequent improvement in practice.

Nascent Research Base: What Do We Know about Teachers' Motivational Responses?

The research conducted on performance management systems provides some information on how these initiatives impact teacher motivation; however, these data are limited in scope. Historically, scholars have documented that performance management policies encounter intense resistance from some teachers, most notably the teachers' unions (Murnane & Cohen, 1986). In 2003, the Public Agenda Foundation conducted a nationally representative survey and found that only 47% of teachers supported financially rewarding those whose students made more academic progress, and further, many teachers in focus groups expressed a visceral reaction to the idea of linking pay with performance (Farkas et al., 2003).

Researchers have documented that teachers react negatively to policies for a variety of reasons – e.g., they do not understand how the policy is designed to operate, they believe policymakers are impugning their level of effort, or they perceive

performance metrics to be unattainable. Research on Florida's performance management initiatives – STAR (Special Teachers are Rewarded) and MAP (Merit Award Program) – discovered how little teachers appeared to even understand how the two recent initiatives operated. Perhaps in part due to their limited understanding, the majority of teachers disagreed that STAR would be able to distinguish between levels of performance (Jacob & Springer, 2007). In the evaluation of the first year of the Texas Educator Excellence Grant (TEEG) program, the majority of teachers (85%) reported that they were already working as hard as they could before TEEG implementation, and as such, only 25% reported that they changed their behaviors as a result of the program (Springer et al., 2008). In another study evaluating the impact of school-based incentives on teacher motivation in Kentucky and Charlotte-Mecklenburg, Kelley, Heneman, and Milanowski (2000), observed that individual teachers' expectation that they could achieve desired outcomes was weaker than initially anticipated.

In contrast, other research has found teachers to be more receptive to changes in performance management. In the evaluation of TEEG, Springer et. al., found that 71% of teachers strongly desired to earn a TEEG bonus and 60% agreed that the TEEG program did a good job of identifying effective teachers. Additionally, more than 90% of the respondents thought increasing student test scores should be of either moderate or high importance in teacher evaluation, making it the highest ranked measure out of 17 indicators (2008). Research has also demonstrated that perceptions among the teacher corps may be changing; indeed, younger teachers are more likely to seek out

opportunities for diverse roles and be in favor of alternate forms of compensation (Blair, 2002; Farkas et al., 2003; Qazilbash, 2007).

As expectancy-value theory would predict, this nascent research base suggests that teachers' attitudes and responses depend on how performance management systems are designed and implemented. In a recent analysis of theories undergirding teacher evaluation systems, Firestone contends that current policies focus primarily on economic approaches to motivation, which emphasize extrinsic incentives (e.g., performance pay, firing ineffective teachers) as opposed to intrinsic approaches, which underscore the importance of building teacher autonomy and support. Though these approaches are not necessarily mutually exclusive, evaluation used for accountability purposes has the potential to undermine the intrinsic incentives that give teachers a sense of control over meeting their own standards of competence (2014). While a growing body of research has begun to examine the impact of recent evaluation systems on student outcomes (Dee & Wyckoff, 2013; Steinberg & Sartain, forthcoming 2014; Taylor & Tyler, 2011), we have limited information on how specific policy design features (e.g., specific measures, observational processes, uses for evaluation) influence teachers' motivation (both extrinsic and intrinsic) to improve their practice.

Expectancy-value theory also suggests that teachers' responses will vary considerably as a function of differences in individual teacher characteristics and school-based organizational factors affecting the process of implementation. Unfortunately, most studies do not take into consideration how new initiatives differentially affect teacher attitudes and subsequently influence motivation and behavioral change (Goldhaber et al.,

2007). Additionally, while there are many studies detailing the importance of school working conditions (Ingersoll, 2001; Ingersoll, 2006; Johnson et al., 2005), existing research does not examine which working conditions motivate teachers in the context of new performance management systems. In the small number of studies where these questions have been investigated, results have not been analyzed within a motivational framework, making it challenging to interpret the divergent findings. Without a deeper understanding of this variation in teachers' motivational responses, system designers do not have enough information to create and implement new performance management initiatives that influence teachers' motivation and subsequent changes in behavior.

CHAPTER 2: METHODS AND DATA COLLECTION

This study fills these gaps in the existing research base by investigating the impact of a new teacher evaluation system in the Aldine ISD, INVEST, on important teacher outcomes. In particular, I investigate several key research questions.

1. What are teachers' attitudes towards the new INVEST system? What are their initial perceptions of the new system's design and implementation?
2. What impact does INVEST have on teachers' motivation and teacher outcomes of interest (i.e., effectiveness and retention)?
    a. Motivation, as measured by teachers' self-reported expectancy and value:
        i. Expectancy. Do teachers believe in their ability to impact their students' progress? Do they believe they will be able to perform well on the system?
        ii. Value. Do teachers value being good at their work? Do teachers value performing well on the new evaluation system?
    b. Effectiveness, as measured by the Aldine Growth Model (a measure of teachers' impact on student growth on standardized exams)
    c. Retention, capturing teachers who left the district at the end of the 2012-2013 year
    d. How is teachers' level of motivation associated with their effectiveness and retention?

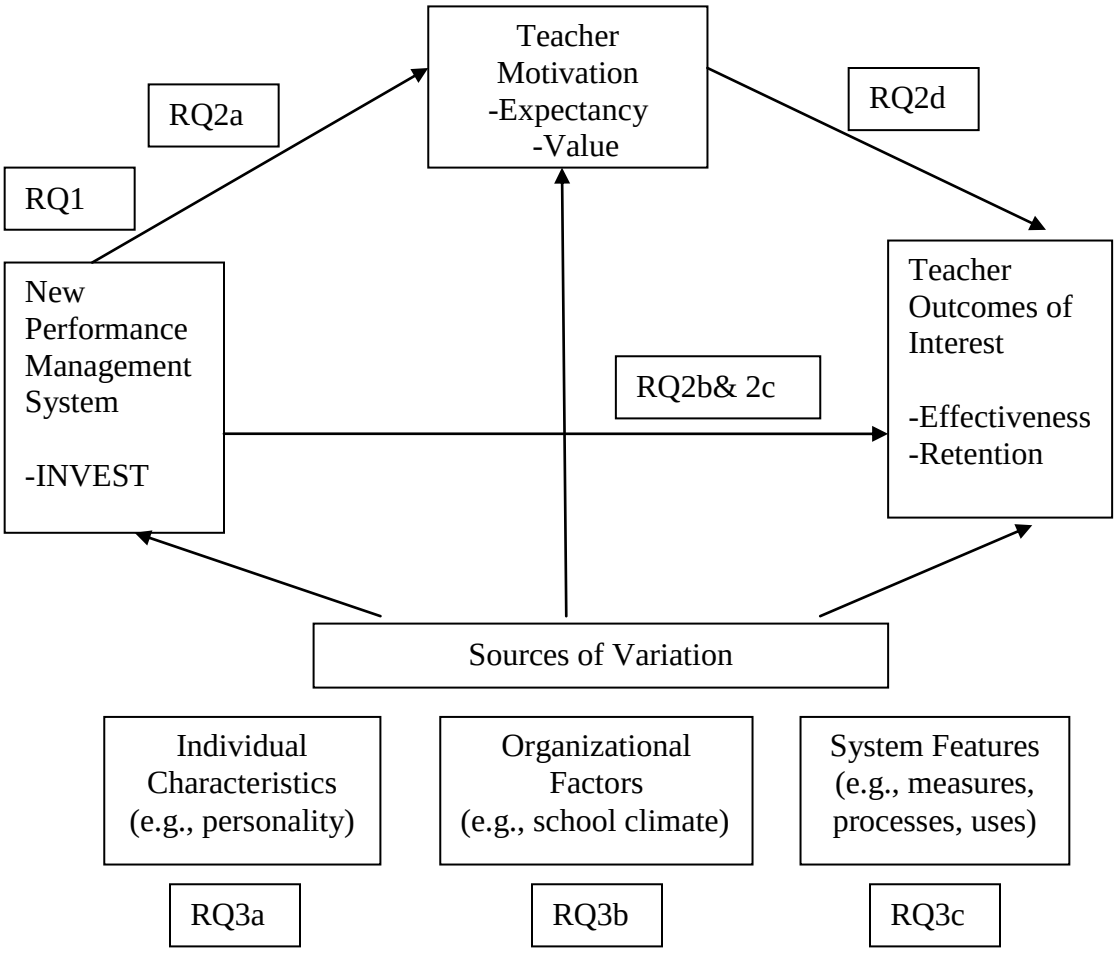

Figure 2-1. Alignment between research questions and motivational framework

3. To what extent are teachers' system motivation, effectiveness, and retention influenced by individual characteristics, school organizational factors, and system features?
    a. Individual characteristics – teachers' personality (i.e., grit, Big 5)
    b. School organizational factors – principal leadership, level of positive support, level of control support, quality of professional community

c. System (design and implementation) features – e.g., perceptions of accuracy and fairness, the quality of feedback, level of understanding of the new system

Methodology

To answer these research questions, I employ a mixed methods design to analyze the impact and implementation of INVEST, a new teacher evaluation system which was piloted in Aldine ISD during the 2012-2013 year. According to Creswell and Clark (2006), mixed methods research focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies. By bringing various perspectives to bear on a policy problem, mixed methods research triangulates data and allows for stronger generalization (Creswell & Clark, 2006).

In this study, I collected quantitative and qualitative data simultaneously, which allowed me to utilize the strengths of both approaches. My quantitative research captures broad-based results from the population of teachers, whereas my qualitative research draws from a small sample and provides a more in depth understanding of how individuals experience policy implementation. For Research Question 1, I used descriptive quantitative data to explore key trends in teachers' responses and supplemented this data with rich qualitative data to understand the rationale and motivation behind teachers' attitudes. For Research Question 2, I relied on quantitative data to examine the overall impact the new system has on teachers' motivation (as captured by survey data), effectiveness (as measured by the student growth measure of the teacher evaluation system), and retention (as reported in administrative data). I

supplemented these data with interview data on teachers' perspectives of the system's impact. For Research Question 3, I again employed mixed methods to understand how outcomes were influenced by system, individual, and school characteristics. Quantitative data provided information on which system, individual, and school factors were most predictive of each of the key outcomes of interest, while qualitative data explored why these factors are so pivotal to teachers' responses to INVEST. In sum, the quantitative indicators provided an overall sense of the new system's impact, while the qualitative data elaborated on why particular effects were observed and how individual and school context shaped responses.

District and System Background

Located in Houston, Texas, Aldine ISD serves an urban population of approximately 64,000 students. More than 84.9% of all Aldine students are classified as economically disadvantaged and receive Title I support, and the racial composition is 70.8% Hispanic, 25.1% African-American. Additionally, 31.9% of the students in Aldine ISD receive support from the Limited English Proficiency or Bilingual programs. Aldine ISD is the recipient of numerous awards including the 2009 nationally recognized Broad Prize for making progress in closing the achievement gap among students of different ethnic groups and socioeconomic statuses and takes a great pride in their approach, which they call the "Aldine way." Rather than relying on outside leadership, Aldine has a home-grown approach to leadership and celebrates its consistent and stable leadership at the administrative level.

Design work on the new teacher evaluation and development system, INVEST, began in September 2011, with the support of Operation Public Education, an external consulting group based at the University of Pennsylvania that I have worked with since 2007. The district used a volume I co-edited, *A Grand Bargain for Education Reform: New Rewards and New Supports for New Accountability* (Hershberg & Robertson-Kraft, 2009), as its guide throughout the design process. This process was inclusive, involving teachers, administrators, and community members. District leadership established three work groups – Teacher Practices, Student Impact, and Other Staff – to work through the many complex decisions required for designing an evaluation system and used the district's democratic process to identify participants for these work groups. Each of Aldine ISD's 74 schools elects five representatives, including two teachers, one paraprofessional, one parent, and one business community member to constitute the Vertical Education Advisory Committee (VEAC). From its members, this group then elects a district-wide body, the District Education Advisory Committee (DEAC). The work groups were composed of VEAC and DEAC volunteers, plus educators with expertise in specific areas (e.g., technology) recruited by administrators. Each work group had between 30 and 60 people depending on the group's purpose, and each met five times over the course of the 2010-2011 school year to design the new system.

The work group recommended specific policy decisions to the district leadership team (which was composed of area superintendents and human resources personnel):

- *Observation.* The district adopted Charlotte Danielson's Framework for Teaching. Originally developed in 1996, the Framework has been used nationally to

document and develop teacher practice. It consists of four broad domains – Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities – further divided into 22 components and a performance rubric that differentiates four levels of performance – Unsatisfactory, Basic, Proficient, and Distinguished.

- *Student Growth.* To measure teacher performance based on student growth, the district decided to use a student growth percentile measure based on the Colorado Growth Model (Betebenner, 2009). The model compares the change in each student's achievement score to all other students in Aldine who had similar achievement scores in the previous year. Each student receives a student growth percentile and the teacher is assigned an overall SGP based on the median SGP of all their students. TAKS/STAAR (the state achievement test in Texas) was used to calculate SGPs in grades 4-9 (and where available in high school subjects), and Stanford/Aprenda was used in grades K-3.
- *Educators Outside of Tested Subjects.* The Danielson rubrics, processes, and protocols were modified to evaluate performance of staff whose work falls outside measures of student growth. The recommendation was made that these educators would also set Student Growth Objectives (SGOs), based on a process pioneered by the Denver Public Schools, to measure their students' progress over the course of the year.

At the end of the year, teachers were rated *Highly Effective, Effective, Needs Improvement*, or *Ineffective* based on meeting pre-determined conditions on each

measure. The "Final INVEST Rating" was drawn from scores on both observation and student growth (either student growth percentiles or student growth objectives). To be *Highly Effective* overall, a teacher must be rated *Highly Effective* in both measures, and to be *Effective*, a teacher must be rated *Effective* in both measures. Teachers will be rated *Needs Improvement* or *Ineffective* if they have received this rating in either of the measures. It is important to note that in the pilot year, the district leadership decided that only the Danielson Framework would be used for consequence (i.e., to put teachers on a professional growth plan) and in the first year, the Student Growth Percentile measure would be reserved for professional development.

INVEST was viewed as a fairly radical departure from the previous appraisal system, Professional Development and Appraisal System (PDAS). Given the significance of the change, district leadership chose to pilot the system so they could incorporate feedback from key stakeholders before rolling out district-wide in 2013-2014. Table 2-1 below depicts the key differences between the current system (PDAS) and the new INVEST system:

Table 2-1

Key Differences between PDAS and INVEST

	Current System (PDAS)	New System (INVEST)
Measures	PDAS evaluates teachers based on principal observation ratings. The ratings are a composite of nine different domains and three different levels of performance. PDAS does not provide a rubric for principals to use when differentiating teacher	INVEST will evaluate teacher performance based on scores on two measures: • *Observation* based on Charlotte Danielson's Framework for Teaching. Each of the 22 components will be accompanied by a

	performance.	detailed rubric that can be used to assess performance. • *Student growth* based on a student growth percentile model (for teachers in tested subjects) and a student growth objectives model (for teachers outside of tested subjects)
Processes	All teachers are evaluated once during the course of the year. Principals conduct walkthroughs but there are no requirements on how frequently these walkthroughs must be conducted. The model is not differentiated based on teacher experience. There are no formal requirements for conferencing between evaluators and teachers.	This model is differentiated to meet the needs of novice and experienced teachers. There will be two tracks – one for novice teachers (in their first three years in the classroom) and one for experienced teachers (more than three years of experience when teachers have received non-probationary status). • *Track 1.* Novice teachers will receive three informal walkthroughs each semester and one formal observation each semester. • *Track 2.* Experienced teachers will receive three informal walkthroughs each year (two in the first semester and one in the second semester). They will receive one formal observation which can occur at any point during the year. All teachers will take part in a goal-setting conference at the beginning of the year and a summative conversation at the end of the year, and each formal observation will be accompanied by both a pre- and post-conference, where evaluators and teachers will discuss progress toward goals.

Uses	Teachers are currently placed on TINAs (Teachers in Need of Assistance) if they are deemed to be underperforming. There are no clear guidelines for why a teacher should be placed on a TINA and in practice, principals use them very infrequently.	Teachers identified in *Needs Improvement* either through walkthroughs or formal observations will be provided with additional support through an individual support plan (ISP) customized to meet their needs. Teachers who continue to not meet standards of practice after four to six weeks will be placed on a professional growth plan (PGP) which will articulate the consequences and disciplinary actions that would occur if performance is not adequately improved. If these goals are not met, teachers will be recommended for non-extension or non-renewal.
Training	Teachers will receive a beginning of the year training in PDAS.	Teachers will receive a beginning of the year training on INVEST. All pilot schools will also receive access to the following professional development resources provided by Teachscape (and aligned to the Danielson Framework): • *The Framework for Teaching Proficiency System*, an online administrator certification process. • *The Framework for Teaching Effectiveness Series*, which is a self-guided, online training system for teachers that features master-scored benchmark videos. • *Reflect Live*, a complete evaluation management system that combines live observation and video-based observation into one platform.

Taylor and Tyler (2011) distinguish between investment in human capital and short-term accountability effects as two possible goals of teacher evaluation policies. In Aldine, both are simultaneously at work. INVEST has several overarching goals:

- *Differentiating and Improving Instructional Practice*. The new evaluation system was designed to differentiate and improve teachers' instructional performance using the Framework for Teaching. Whereas in 2010-2011, 96% of teachers were simply rated "satisfactory," one of the goals of this new system was to increase dialogue about improving practice and provide a more accurate picture of teacher performance across the district's schools.
- *Increasing the proportion of highly effective and effective teachers.* To raise the quality of the district's teaching force, another goal of the new system was to increase teacher effectiveness. This growth will be accomplished by identifying teachers in need of improvement, providing targeting support, and dismissing those who are unable to improve the quality of instruction.
- *Reducing teacher retention (of high performers).* The final system's goal was to increase teacher satisfaction and thus reduce the rate of teachers who leave the Aldine ISD, particularly among highly effective educators.

Sample

In spring 2012, Aldine ISD strategically selected 34 of the district's 74 schools to participate in the Year 1 pilot of INVEST. The goal was to ensure that the selected schools were as representative of the district schools as possible to learn how the initiative would work in a variety of settings. To accomplish this goal, district leadership

strategically selected schools that varied along a number of dimensions – i.e., level (elementary, middle, high), student performance level (on both achievement and growth measures), demographics (percent LEP, percent economically disadvantaged). Though the pilot schools were not randomly selected, there were no statistically significant differences between the pilot and control schools on key baseline measures. All of the schools in the AISD are Title I, meaning they have a significant percentage of students who are low-income and on free and reduced priced lunch. Additionally, the district is composed almost entirely of minority students, though there is variation in the percentage of African-American students and Hispanics across campuses. During the 2012-2013 year, there were 4,397 teachers teaching in these 74 schools and 1,883 or 43% of these teachers were in pilot schools. This sample includes teachers outside of traditional subjects (e.g., art, music), as well as other staff (e.g., counselors, nurses).

From the 34 pilot schools, I identified six schools for in-depth qualitative data collection. The sampling strategy was used to capture variation across levels (e.g., elementary, high) and school performance levels (e.g., both higher-performing and lower-performing schools). The goal was to create an overall case study sample that was as diverse as possible, representing different school environments. The school selection process is summarized in the Table 2-2 below.

Table 2-2

School Selection Process

Level	Lower Performing	Higher Performing
Elementary	X	X
Intermediate	X	X
High School	X	X

Quantitative Methods and Analysis

Measures

Teachers provide critical information on the rollout of implementation efforts and the new initiative's impact on their effort and attitudes. As such, a major source of data for this study was a teacher survey I administered to the population of teachers in Aldine ISD in both pilot and non-pilot schools. This survey provided critical information at the beginning of the year that I compared with information at the end of the year to assess the impact of the pilot on teacher motivation. It also provided critical information on how the impact of the pilot was influenced by characteristics of both individual teachers and schools.

Survey questions fell into one of several categories: (1) teacher motivation, (2) individual teacher characteristics, (3) school working conditions, and (4) attitudes toward teacher evaluation. At the beginning of the year, the survey included questions on teacher motivation, individual personality characteristics, school working conditions, and a few questions on teachers' attitudes toward evaluation. Since teachers had not yet experienced the new evaluation system, these questions asked for perceptions of evaluation in more

general terms. At the end of the year, the survey included the same questions on teacher motivation and school working conditions, as well as a more extensive set of questions on attitudes toward teacher evaluation and specific questions on the new INVEST system (for teachers in pilot schools). Since personality characteristics are relatively stable, these questions were not included on the end of year survey. A more detailed description of measures is included in Table 2-3. I modified several of these measures – i.e., expectancy, value, the Big 5, grit, administrative leadership, control, support, and professional community –from pre-existing scales. Table 2-3 also reports the Cronbach's Alpha associated with the relevant scales from this survey administration.

Table 2-3

Survey Measures Used in Analysis

Measure	**Survey Item**	**Cronbach's Alpha**
Individual Personality Characteristics		
Teaching Grit	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). • Right now, my interest in teaching is about the same as it was before the school year began • I am working as hard as I did at the beginning of the school year • Lately, setbacks have not discouraged me • Every day, I actively try to improve my teaching • At the moment, nothing is more important to me than improving my teaching • In my work, I always persevere, even when things do not go well	.75

Overall Grit	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). *I see myself as someone who:* • Is not discouraged by setbacks • Finishes whatever I begin • Is diligent and an extremely hard worker • Had been obsessed with a project for a short time but later loses interest • Often sets a goal but later chooses to pursue a different one • Has difficulty maintaining focus on projects that take more than a few months to complete	.68
Conscientiousness	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). *I see myself as someone who:* • Does a thorough job • Does things efficiently • Tends to be lazy	.59
Extraversion	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). *I see myself as someone who:* • Is talkative • Is outgoing, sociable • Is reserved	.71
Agreeableness	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). *I see myself as someone who:* • Has a forgiving nature • Is considerate and kind to almost everyone • Is sometimes rude to others	.59

Emotional Stability	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). *I see myself as someone who is:* • Worries a lot • Relaxed, handles stress well • Gets nervous easily **Note: this scale was reverse coded for ease of comparison*	.60
Openness	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). *I see myself as someone who is:* • Is original, comes up with new ideas. • Has an active imagination • Values artistic experiences	.65

School Working Conditions

Quality of administration	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • The administration's behavior toward staff is supportive and encouraging • My principal enforces school rules for student conduct • The principal knows what kind of school he or she wants and has communicated that vision	.83
Positive support	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • I receive a great deal of support from parents for the work that I do • Necessary materials are made available • I am given the support I need for students	.57

	with special needs	
Level of control	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • I have control over selecting content, topics, and skills taught in my classroom • I have control over selecting teaching techniques • I have control over disciplining students	.59
Presence of a Professional community	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • Rules for student behavior are consistently enforced by teachers • There is a great deal of cooperative effort among staff members • Most of my colleagues share my beliefs about the central mission of the school	.68
Teacher Evaluation Attitudes		
Quality of Evaluation Measures	How Much Do You Agree that the Evaluation Measures Were (strongly disagree; disagree; neutral; agree; strongly agree): • Specific and clear • Accurate and fair • Comprehensive • Student-centered	.91
Fairness of Evaluation Process	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • Overall the evaluation system was fair • The observation accurately captured my performance • I agree with my evaluator's assessment of my performance	.90

Frequency of Evaluation	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • My evaluator spent adequate time this year observing me • My evaluator spend adequate time meeting with me to discuss my practice Number of observations and number of conversations	.89
Quality of Feedback and Growth	How Much Do You Agree With the Following Statements About the Teacher Evaluation System (strongly disagree; disagree; neutral; agree; strongly agree): • Encouraged my professional growth • Provided feedback that identified specific areas for improvement • Resulted in changes in my practice	.84
Teacher Perceptions of INVEST (Pilot Schools Only)		
Level of Understanding	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • The information I received about INVEST at the beginning of the year provided me with an understanding of the new evaluation system • The information I received about INVEST throughout the year improved my understanding of the new evaluation system • The Teachscape modules provided me with an understanding of the Danielson component of the new evaluation system • The Student Growth percentile modules provided me with an understanding of the SGP component of the new evaluation system.	.84
Positive Goal-setting	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree):	.76

	• The goal-setting/action planning process at the beginning of the year helped me focus my goals for improving my teaching performance • This year, because of INVEST, I set more challenging goals for myself than in previous years	

Accuracy of INVEST Measures	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree):
	
	• Overall the Danielson Framework measure used to evaluate my teacher performance under INVEST provides an accurate and comprehensive picture of my teaching.
	  o Domain 1 (Planning and Preparation) is accurate and fair
	  o Domain 2 (Classroom Environment) is accurate and fair
	  o Domain 3 (Classroom Instruction) is accurate and fair
	  o Domain 4 (Professional Responsibilities) is accurate and fair
	• Student Growth Percentiles are an accurate and fair measure of my teaching performance
		.92

Positive Impact of INVEST	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree): • INVEST provides specific feedback on areas to improve my teaching • INVEST provides the support I need to improve my teaching • INVEST will help me improve my teaching • INVEST will support teacher development • INVEST will lead to improvements in student growth and achievement	.93

Teacher Outcomes

Motivation

Personal Expectancy (belief in ability)	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree). • I can get through to the most difficult students • I can promote learning when there is a lack of support from home • I can motivate students who seem to have lost interest in school work	.74
Personal Value (value for work)	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree) • Compared to my other roles in life (e.g., parent, friend, community member), it is important for me to be an effective teacher • In general, I find teaching to be interesting work • I enjoy being a teacher	.66
System Expectancy (belief in ability on INVEST system)	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree) • It is possible to reach the Highly Effective level on the new INVEST system	
System Value (value for INVEST system)	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree) • I want to be considered Highly Effective on the new INVEST system	
Changes in Practice	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree) • I implemented changes in my practice as a result of the new evaluation system	

Retention	
Teacher Burnout	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree) • I feel emotionally drained from my work • I feel fatigued when I get up in the morning and have to face another day • I feel frustrated by teaching
Teacher Turnover Intentions	How Much Do You Agree With the Following Statements (strongly disagree; disagree; neutral; agree; strongly agree) • I will probably look for a new job in the near future • At the present time, I am actively searching for another job • I do not intend to leave teaching at my school

In addition to this survey data on teachers, I used longitudinal administrative data collected by the district from the 2010-2011, 2011-2012, and 2012-2013 school years on teacher effectiveness and retention. These measures are captured in Table 2-4.

Table 2-4

Administrative Data

Effectiveness	
Observation (Danielson) – pilot schools only	Teachers' score on the Danielson Framework for Teaching • Average score over four components of the Danielson Framework (on a scale of 1-4) • Teachers' overall rating (Ineffective, Needs Improvement, Effective, Highly Effective)

Student Growth Percentile – teachers in tested subjects	Teachers' score on the Student Growth Percentile • Teachers' median student growth percentile for their class (on a scale of 1-100) • Teachers' overall rating (Ineffective, Needs Improvement, Effective, Highly Effective)
Retention	
Teacher Retention	Teachers' retention • School-level aggregate teacher turnover rate (available for 2010-2011, 2011-2012, and 2012-2013 school year) • Teacher-level turnover (only available for 2012-2013 school year) – whether the teacher stayed teaching in the district

Administrative data also provided information on school demographics, such as ethnicity (percent African-American and Hispanic), free and reduced price lunch status (a proxy for poverty), and the percentage of Limited English Proficient (LEP) students. As demonstrated in Table 2-5, these data were used to ensure that student and teacher covariates were balanced across pilot and non-pilot schools. For student covariates, non-pilot schools had a slightly higher percentage of LEP students, though this difference was not statistically significant. There were no statistically significant differences between pilot and non-pilot schools in terms of ethnicity or the proportion of students who qualified for free and reduced priced lunch (i.e., low income), nor were the differences between pilot and non-pilot schools' student growth (aggregated at the school level) from the previous school year (2011-2012) significant. For teacher covariates, the pilot and non-pilot schools also appeared to be fairly balanced, which is important since the

Table 2-5

Comparison of Pilot and Non-Pilot School Characteristics

Variable	Pilot Schools (*N*=34) *M*	*SD*	Non-Pilot Schools (*N*=40) *M*	*SD*	*p-value*
Student Growth*					
Reading	49.31	7.13	48.50	8.86	.50
Math	48.62	11.06	48.29	11.74	.84
Student Demographics					
African-American	27.3%	20.28	27.6%	16.34	.94
Hispanic	68.6%	21.05	67.9%	17.02	.88
Low-income	85.1%	8.36	86.1%	7.32	.61
Limited English Proficient	31.3%	22.66	34.4%	23.02	.57
Teacher Demographics					
Ethnicity (white)	34.2%		36.0%		.26
Gender (female)*	79.7%		76.9%		.03
Certification (traditional)	58.3%		57.5%		.69
Average years	10.51	2.26	9.69	2.35	.14
First five years	40.8%	13.75	44.3%	14.22	.28
Turnover 2012	9.75	4.30	10.26	4.64	.63

Note. Student growth data only exists for school with tested subjects, *N* = 29 for pilot schools, and *N* = 34 for non-pilot schools.

intervention targeted teacher practice. There were slightly higher percentages of white and male teachers in non-pilot schools compared to pilot schools, but only the gender difference was statistically significant. Since the pilot differentiated support along years of teaching experience,

it is important to note that though pilot teachers are slightly more experienced than non-pilot teachers, these differences were not statistically significant.

I also examined covariate balance on measures of school working conditions and initial perceptions of teacher evaluation. As demonstrated in Table 2-6, none of the differences between pilot and non-pilot teachers' attitudes towards working conditions and perceptions of evaluation was significant. Across the board, pilot schools appeared to score slightly higher on measures of school climate, though these differences were not statistically significant. Teachers in pilot schools had slightly lower beginning of the year perceptions of evaluation measures, as well as attitudes towards the fairness and supportiveness of the process. This could be a function of the fact that teachers in pilot schools were aware of the fact that their evaluation system was changing and had received an initial introduction to INVEST at the time of survey administration. Nonetheless, these differences were not statistically significant.

Table 2-6

Comparison of Pilot and Non-Pilot Schools School Climate at Baseline

	Pilot Schools (N=34)		Non-Pilot Schools (N=40)		
Variable	*M*	*SD*	*M*	*SD*	*p-value*
Climate					
Administration	4.07	0.28	4.01	0.27	.36
Support	3.42	0.25	3.39	0.25	.51
Professional Community	3.81	0.25	3.76	0.27	.42
Control	3.65	0.17	3.63	0.20	.52
Perceptions of Evaluation					
Growth	2.74	0.28	2.81	0.26	.28
Observation	3.35	0.24	3.41	0.26	.34
Fairness	3.74	0.17	3.80	0.19	.20
Professional Growth	3.64	0.25	3.73	0.21	.08

*$p < .05$. **$p < .01$. ***$p < .001$

In any intervention analysis, the concern is that the sample participating in the pilot may have different characteristics from the population as a whole and that any observed treatment effect will be incorrectly attributed to the intervention. These analyses suggest that though the pilot schools were not chosen at random, they are fairly representative of the district as a whole on student characteristics, teacher demographics, and school climate indicators. While we cannot conclude that they were not substantively different on unobservable characteristics, this baseline equivalence strengthens the inference we can draw from the impact analysis.

Procedures

In the summer of 2012, I shared an initial draft of the teacher survey with district leadership for feedback. After making minor modifications, I piloted the survey with approximately 30 teachers in the Philadelphia region. This piloting process ensured that questions were phrased clearly and captured sufficient variation in teacher responses. To ensure the survey was a minimal administrative burden and protected teachers' confidentiality, I created a cover page accompanying each survey that assigned each teacher a unique teacher ID, which I then matched with the district database. Upon receipt of the survey, teachers could remove the cover page with identifying information and keep for their records such that all survey results will be deidentified moving forward. Teachers were provided with an overview of the project and an informed consent letter, both of which were approved by the University of Pennsylvania Institutional Review Board.

Table 2-7

Comparison of Respondents Completing Both Surveys and Non-Respondents

	Respondents (*N*=2662)		Non-Respondents (*N*=1735)	
Variable	*M*	*SD*	*M*	*SD*
Ethnicity*				
Percent White	36.3%		33.5%	
Percent Hispanic	24.3%		22.8%	
Percent AA	34.6%		39.7%	
Asian	2.8%		2.0%	
Gender				
Male	22.3%		21.3%	
Female	77.7%		78.7%	
Experience				
Years in district	8.14	7.14	8.13	7.24
Years in teaching	11.28	8.86	11.20	8.78
Performance				
Observations	3.19	.33	3.22	.44
Student Growth *	51.61	13.07	48.64	13.46

Note. *N* = 1652 for Observations and *N* = 906 for Student Growth
*$p < .05$. **$p < .01$. ***$p < .001$

At the end of August, Aldine ISD principals administered the finalized beginning of the year survey I developed to their teachers during a campus professional development. In total, 3647 surveys were completed, out of a population of 4178 teachers, for a response rate of 84%. At the end of May, principals administered the end of year surveys I developed to the same population of teachers in addition to 219 new

hires to the district (for a total sample size of 4397 teachers), and 3254 surveys were completed for a response rate of 74%. In total, 2662 teachers completed both the beginning and end of year samples for an overall response rate of 61%. Of the 2662 teachers, 59% (or 1565) were in control schools and 41% (or 1097) are in pilot schools.

As demonstrated in Table 2-7, respondents who completed both surveys were more likely to be White and when in tested subjects, were more likely to perform well on the student growth measure, than teachers who did not respond to the survey. However, though these differences are statistically significant, they are relatively small in magnitude. There were no significant differences on any other demographic or performance indicators, suggesting that respondents are fairly representative of the population of teachers.

Analysis

I used responses from these surveys to assess teachers' attitudes toward the new evaluation system, as well as to investigate how their motivation and performance were influenced by individual characteristics and perceptions of school-based organizational factors. To answer Research Question 1, I summarized the level and distribution of responses to each survey question and compared results across different types of schools (e.g., high versus low performing) and types of teachers (e.g., novice versus experienced, effective versus ineffective). After assessing the reliability of the motivation, personality, and school climate scales (presented in Table 2-3 above), I used exploratory factor analysis to determine how the questions on teachers' attitudes toward evaluation could be reduced to a smaller number of components. Using the Kaiser criterion, I kept any factor

with a corresponding eigenvalue greater than 1 and then created factor scores representing each individual's placement on the factor that could be used in subsequent analyses.

Following this descriptive analysis, I assessed the impact of INVEST on teacher motivation, teacher effectiveness, and teacher retention (Research Question 2) using a quasi-experimental technique called difference in differences (DID). To examine the impact of a treatment, DID presumes that we must compare the treatment group after treatment both to the treatment group before treatment and to some other control group. In this study, the treatment group was those schools piloting the INVEST system, while the control group was those implementing the traditional teacher evaluation system. Subtracting the pre-treatment difference in outcomes from the post-treatment difference eliminates one kind of selection bias, namely the kind related to time-invariant individual characteristics. In other words, if what differentiates pilot and non-pilot schools is fixed in time and any changes are identical between the two groups, subtracting the pre-treatment differences eliminates selection bias and produces a plausible estimate of the impact of the INVEST initiative.

A causal interpretation of the difference-in-differences estimator rests on one — untestable — assumption: that in the absence of the policy the pilot schools would have continued to have the same rate of change in the outcome variable (i.e., teacher motivation, effectiveness, retention) as the control schools. One way to examine this assumption is to examine pre-treatment trends between pilot and non-pilot schools for the outcome of interest. As demonstrated in Figure 2-2, the pilot schools had a lower

turnover rate than non-pilot schools at the end of the 2011 school year (8.04% compared to 9.44%), but this percentage increased at a slightly faster rate during the 2011-2012 year (the year prior to the pilot) in pilot schools (+1.71 compared to .82). This provides some evidence that for the retention outcome, the difference-in-differences assumptions may not hold. Teacher motivation and effectiveness data were only available for the year prior to the pilot, so unfortunately this analysis could not be conducted for these outcomes.

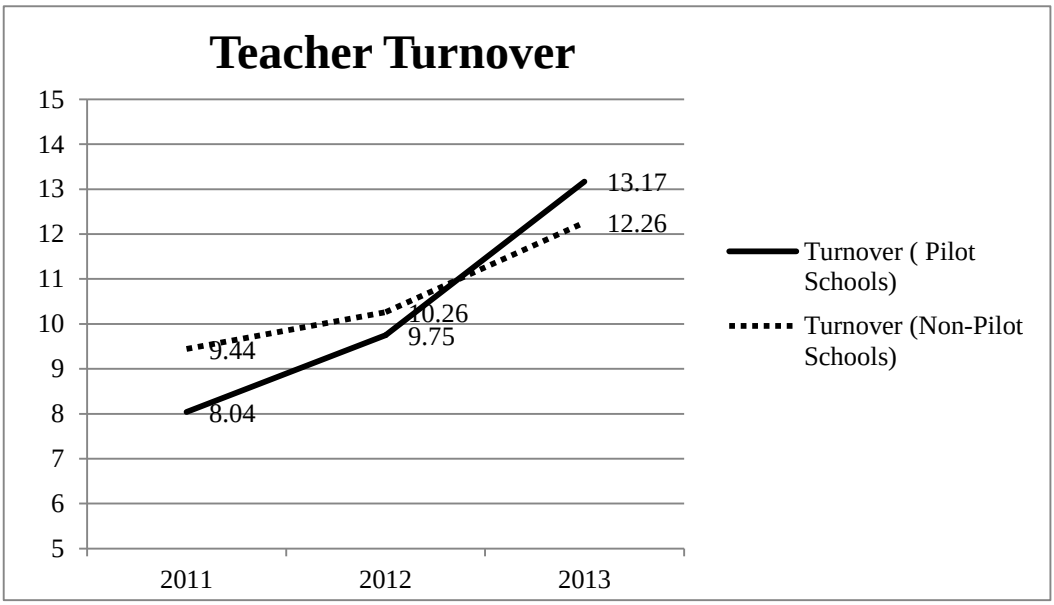

Figure 2-2. Percentage of teacher turnover over time

Note: This figure represents the percentage of teacher turnover at the end of each year (2010-2011, 2011-2012, 2012-2013**).** The pilot was implemented in the 2012-2013 school year.

To attempt to account for differences between the initial composition of treatment and control groups that may influence this rate of change, I ran my analyses with and

without school fixed effects and controlled for teacher characteristics (e.g., years of experience). However, I am still unable to account for time-varying unobservable characteristics. For the causal interpretation to hold, these time-varying characteristics must affect the pilot and non-pilot schools in the same way.

The basic difference-in-differences model takes the following form:

$$Y = \beta_0 + \beta_1 * T + \beta_2 * P + \beta_3 * (T * P) + \Gamma * X + \varepsilon$$

Y represents the outcome variable of interest in each set of schools over the course of the 2012-2013 school year – and teacher motivation (operationalized by survey questions on expectancy and value) and teacher effectiveness and retention (using administrative records). *T* is a time dummy, *P* is a pilot dummy, and *T***P* is the interaction of the time dummy and the pilot dummy. β_0 is the baseline average for the non-pilot schools, β_1 represents the change in outcomes over the year in the control group, β_2 represents the differences between the pilot and non-pilot schools before the implementation of INVEST, and β_3 represents the impact of INVEST. X is a vector of covariates that may affect outcomes (e.g., student demographics, school performance, leadership quality, teacher demographics) and Γ is the coefficient associated with these covariates. The approach is further explicated in Table 2-8 below:

Table 2-8

Difference-in-Differences Approach

Outcomes	Non-Pilot Schools	Pilot Schools
Pre-INVEST	A	B
Post-INVEST	C	D
Coefficient	Calculation	
β_0	A	
β_1	C – A	
β_2	B – A	
β_3	(D-B) – (C-A)	

To answer Research Question 3, I explored variation in teachers' responses to the policy. First, I used multiple regression analyses to evaluate how teachers' individual characteristics (i.e., grit, Big 5, experience), school-based organizational factors (i.e., school climate, leadership), and attitudes towards system features (e.g., perceptions of accuracy and fairness, quality of feedback and growth) predicted the three outcome variables of interest – teacher motivation, teacher effectiveness and retention. I began with a basic model controlling for demographic characteristics and added in sets of predictors to assess the additional predictive power of various types of factors – i.e., individual characteristics, school characteristics, and system characteristics.

Qualitative Data: Methods and Analysis

Though this quantitative analysis supplies data on the impact of INVEST in Aldine ISD, it does not provide a fine-grained analysis of how teachers experienced the new policy. To gather more in-depth information on how the pilot impacted teachers'

motivational responses, I conducted qualitative research in a subset of six pilot schools. These data were used to supplement the more comprehensive information from the teacher survey. At each of these six schools, I interviewed the administrator and six teachers, selected purposively to vary across performance levels (i.e., effectiveness levels based on SGP data from 2011-2012) and experience levels (i.e., novice vs. experienced teachers). See Table 2-9 below for a demonstration of how teachers were chosen for participation in the study.

Table 2-9

Teacher Selection

Performance Level	Novice	Experienced
Ineffective	X	X
Effective	X	X
Highly Effective	X	X

I interviewed administrators and participating teachers at the end of the first semester of implementation (late November/early December) and the end of the year (May) to capture feedback at various stages of the implementation process.

- Round 1. In late November/early December 2012, I conducted interviews with administrators and teachers in the six case study schools to capture initial feedback on the new teacher evaluation system.
- Round 2. In early May 2013, I conducted the final round of interviews with administrators and teachers in case study schools, to capture feedback after teachers had received their end of year review.

Each interview lasted approximately 30 minutes. During these interviews, I gathered information on teachers' and administrators' perceptions of the new evaluation system and its impact: specifically: (1) questions related to the value teachers and administrators placed on the new measures; (2) questions related to the perceived impact the new system would have/was having on teacher motivation, behavior, and performance; and (3) factors affecting implementation of the new system. All interview protocols were grounded in the research questions but also included open-ended questions to allow interviewees to guide the conversation. All protocols were shared with district leadership for feedback and then piloted before being used in actual case study settings. The piloting process ensured that questions were phrased clearly and able to gather the desired information.

I also reviewed district documents and attended monthly meetings of the leadership team over the design and pilot school years (2010-2013). These meetings were used to collect additional information on the goals and design process undergirding the new evaluation system. Another purpose was to document district leaders' experiences implementing the new evaluation system, by identifying which aspects were challenging and how the district addressed those challenges, as well as which factors affected the success of the implementation roll-out.

After conducting this data collection, I generated three data sources from the interviews: interview notes, interview transcripts, and memos. I drafted memos following each visit and included initial impressions from the interviews regarding key issues such

as school culture, themes across teacher reactions, and/or interactions with staff. Finally, there were digital recordings for interviewees who consented to be audiotaped.

To help ensure interviewees felt comfortable being candid about their perspectives on the new system, I assured them that neither their names nor the names of their schools would be revealed in any official report. Interviewees were also informed that their responses would be aggregated with others in the school and district to get an overall picture of INVEST. All interviewees were given detailed consent forms which had been approved by the University of Pennsylvania Institutional Review Board. To protect the confidentiality of interview data, I stored data, including recordings and transcripts, on a password protected server and removed identifiers from all analysis.

To aggregate information from interviews, I used Atlas.ti qualitative software to create a coding scheme for interview transcripts that included both inductive and deductive codes. I applied this coding scheme to create a case study of each school in the analysis. These case studies mirrored the questions in the interview protocols and systematically examined how each school implemented the new INVEST system and how teachers responded to the key features of the new system. After completing an individual case study for each school, I investigated how implementation varied across different types of schools (i.e., by level, performance) and how school-level characteristics (e.g., leadership, professional community) contributed to this variation.

After completing case studies for each of the six schools, I analyzed the coded transcripts for trends in responses across teachers. Using Atlas.ti, I created codes that captured teachers' responses to INVEST, as well as their individual personality

characteristics (i.e., grit, Big 5). I assessed each individual teacher across each of the codes and used this data to create five teacher profiles which categorized teachers' responses to INVEST. Each profile was assigned a name that described their reaction – the invested teacher, the sponge teacher, the burnt-out teacher, the insulted teacher, and the skeptical teacher. Two research assistants working on the project also reviewed the data and confirmed the placement of each teacher, corroborating the usefulness of the profile categorization. Data collection methods are summarized in Table 2-10 below.

Table 2-10

Data Collection

	Sample Size			Collection Schedule			
Measure	Pilot Schools	Non-Pilot Schools	Total *N*	Summer/ Fall 2012	Winter 2013	Spring 2013	Summer/ Fall 2013
Teacher survey	34	40	*N* = 4397	X		X	
Administrator/ teacher interviews	6	--	*N* = 42		X	X	
Student records	Student achievement and demographic data for all students			X			X
Employee records	Administrative data for all teachers and principals			X			X
Performance evaluation system results	Evaluation system data for all teachers						X

PART ONE FINDINGS: OVERALL

INVEST was piloted in 34 of Aldine ISD's 74 schools during the 2012-2013 school year following an intensive year of work group meetings, which involved teachers and administrators in the design of the new system. For teachers in pilot schools, INVEST replaced the previous evaluation system, the Professional Development and Appraisal System (PDAS) and evaluated teachers on two measures of teaching performance, the Danielson Framework (observation) and Student Growth Percentiles (student growth). However, during the pilot year, only the observation measure was used for accountability purposes (i.e., to place struggling teachers on improvement plans). The system was differentiated to meet the needs of new and experienced teachers, with additional observations and conversations for novices. To support rigorous implementation, principals were required to pass a certification exam on the new Danielson Framework using an external process provided by Teachscape. All teachers viewed the same videos as administrators and then took part in a goal-setting process, where they reflected on their practice and set performance goals for the year.

This first part of the dissertation draws on both quantitative and qualitative data to provide an overview of the overall trends gathered on system implementation and impact. I use survey data to compare the experience of teachers in pilot schools who completed the beginning and end of year surveys (N = 1097) with teachers in non-pilot schools remaining under the traditional PDAS system who also completed both surveys (N = 1565). This data was supplemented by qualitative teacher interview data collected in

pilot schools (N = 36) as well as informal interviews and meetings with the district leadership team. The results are divided into two sections.

- ***Chapter 3: System Implementation Descriptive Analysis.*** This chapter answers Research Question 1, by examining overall trends in teachers' attitudes towards INVEST. I use both quantitative and qualitative data to describe how teachers experienced the pilot year of implementation.
- ***Chapter 4: Overall System Impact.*** After presenting descriptive results, this chapter investigates the impact of INVEST on teacher motivation, effectiveness, and retention (Research Question 2). I use the difference-in-differences approach to estimate the pilot's impact on each of these outcomes and then examine the qualitative data to better understand how teachers' attitudes translated into these results.

CHAPTER 3: SYSTEM IMPLEMENTATION DESCRIPTIVE ANALYSIS

After administering a beginning of the year survey to establish baseline equivalence in fall 2012, I gathered data in two phases: winter 2012 and spring 2013. In Phase 1 (November-December), I collected qualitative data on early implementation of the new system through interviews with teachers and administrators in six case study schools. In Phase 2 (May), principals administered a confidential end of year survey I developed to capture information on teachers' attitudes towards specific aspects of the new system. During this phase, I also revisited the same case study schools to gather data on how teachers' perceptions had changed over the course of the school year. This chapter provides an overview of the key descriptive data on system implementation, by exploring overall trends, as well as investigating how these overall trends varied based on subgroups of teachers and schools. Accordingly, it is divided into two sections:

- ***Section 1: Overall Trends.*** This section provides an overview of the key descriptive results (both quantitative and qualitative) from the two phases of data collection. It highlights overall perceptions of evaluation and explores how these attitudes changed over the course of the year.
- ***Section 2: Subgroup Analysis.*** This section explores variation in teachers' responses to the new system across specific subgroups of teachers and schools. It uses quantitative survey data and qualitative interview data to investigate how perceptions varied across subgroups of teachers (i.e., experience, effectiveness) as well across types of schools (i.e., school level, school performance).

Section One: Overall Trends

Phase 1: Mid-Year

When I first visited schools in November and December, INVEST was still in the early months of its first year of implementation. As may be expected with the roll-out of any new system, many principals had struggled to consistently execute INVEST's increased requirements, in particular the additional observations under the new system. In the words of one principal, INVEST was a *"complete shift from PDAS [the old system]*" which made it *"a heck of a lot of work"* (School 5, Principal). All of the principals I interviewed noted the considerable time they were spending on each teacher observation compared to previous years, due to additional expectations around detailed scripting of the lesson and logging results into the Teachscape technology platform. The increased time demands, particularly as they were learning the new system, made it challenging for many of the pilot principals to maintain their schedule for evaluations. Consequently, several of the teachers I interviewed mid-year had yet to be observed or receive feedback on their instruction. During interviews, rather than report on their experiences with actual implementation, these teachers instead shared their anticipated expectations. Though responses varied, several trends emerged as consistently influencing teachers' attitudes towards the new system in these early months of implementation – level of understanding of the purpose of the new system, attitudes toward system accuracy and fairness, and opinions on the quality of feedback and opportunities for professional growth.

Understanding/Purpose. Prior to the launch of the pilot, the district leadership created a centralized handbook and PowerPoint explicating the features of INVEST. These materials focused on the need for change and provided a description of the new evaluation measures, in an attempt to build teachers' understanding of – and investment in – the new system. In particular, the INVEST brochure (developed explicitly for teacher communication) emphasized the importance of supporting teacher development and advancing high expectations for both students and educators. When compared to the prior Professional Development and Appraisal (PDAS) system, the brochure stated that *"the new system (INVEST) will foster professional conversation, provide more thorough observations, and give teachers the opportunity for growth"* (INVEST teacher brochure). During the week prior to the start of the school year, principals were expected to share this information on the purpose and design of the new system with their staff during orientation sessions.

Even with the existence of these centrally developed resources, principals' presentation to their teachers on the purpose of the new system varied considerably. As a result, at the beginning of the year, teachers initially had two very different understandings of the purpose of INVEST – there were those who believed the system would result in improved teaching and learning and those who believed the system was designed primarily as a tool to hold teachers accountable for their performance. Though there was some overlap between the categories (where teachers believed the system could realize both goals), the majority of teachers I interviewed appeared to either view the system as designed for one purpose or the other. Teachers who believed the system was

intended to support professional growth shared that INVEST was a tool to support teachers' development: *"The purpose of INVEST is to see exactly where our strengths are, what we can do to build on those, and what are weaknesses are. It helps make us into the best teacher we can be"* (School 1, Teacher 3). In contrast, other teachers shared that INVEST initially increased teachers' anxiety as it was *"just another way to make the teachers accountable.*" To intensify these fears, some teachers reported hearing rumors that INVEST was devised to make it easier for leadership to not renew contracts given budgetary challenges at the state level: *"Like most people in the teaching profession now, I was thinking it is a tool to get rid of teachers or make it harder for them to achieve high standards"* (School 5, Teacher 3).

Differences in teachers' responses appeared to be associated with the district's decentralized communication strategy. Though resources had been developed at the district-wide level, the end of year survey revealed that only 15% of teachers in pilot schools reported consistently accessing the district's online portal or website for information on INVEST. Instead, teachers primarily relied on their principals to provide information on the purpose and expectations of the new system. Though there was considerable variation in the quality of principal communication across schools (which will be discussed in more detail in Chapter 5), as demonstrated in Table 3-2, overall only 54% of the teachers in pilot schools reported receiving information at the beginning of the year that provided them with an understanding of the new evaluation system.

In an attempt to build understanding, district leadership had required teachers to watch a series of modules on the Danielson Framework (the same Teachscape modules

that administrators watched during their certification process) that lasted 16 hours. Though these modules were intended to invest teachers in the new system by providing them with detailed information on system expectations, for many teachers, they had the adverse effect. One teacher shared how the workload heightened frustration and led teachers to believe the system was focused on accountability: *"It's just so much extra work. This is just ridiculous is the word I keep hearing. We're already doing so much as it is and then they're like, do all this on top of it [referring to the modules] because we want to evaluate you, which is unfair"* (School 5, Teacher 2). Indeed, across the board, teachers and administrators believed that the expectations at the beginning of the year were too demanding and the timeline was rushed, which made the introduction of the new system quite overwhelming. The majority of teachers complained that INVEST had increased expectations without providing additional time to meet those expectations or reducing other responsibilities.

Unlike the Danielson Framework, teachers had not received substantive training on Student Growth Percentiles (SGP measure) by November/December, so many also raised questions about how student growth would factor into their overall evaluation. These questions varied considerably, but most commonly were concerned with the rigor of the new state-mandated assessment and how the metric could be expected to account for the fact that students had such significantly different starting points. Unlike with the Danielson Framework, most teachers' questions were hypothetical, as they still knew very little about how the SGP measure would work in practice.

Accuracy/Fairness. Despite their frustration with the increased expectations for the workload under the new system, the majority of teachers and principals found the Danielson Framework to be an accurate and fair measure of teaching performance. According to teachers' perspectives, the Framework was comprehensive, specific and student-centered, all of which contributed to initial positive perceptions. As one teacher noted, the comprehensive nature of the Framework meant the rubric captured her daily performance as a teacher, *"It really allows you to see what a teacher should be doing every single day… Those four domains really capture what a teacher does"* (School 1, Teacher 3). Many teachers were especially appreciative of the specificity of the Framework, because it meant they knew exactly what was expected of their performance: *"It's black and white. You can really see what they're looking for …and know exactly what actions are expected for each component"* (School 4, Teacher 4). Additionally, teachers believed that unlike PDAS, the Framework challenged them to create student-centered classrooms and empower their students as learners. As one teacher remarked, *"I like the fact that it is more centered on the students. To earn 4s, you have to get the students generating the conversation… you know, it's forcing the teachers to become facilitators and empowering student"* (School 1, Teacher 6).

In addition to appreciating the observation measure, teachers also shared positive perceptions of the observation process itself. Under the new INVEST system, teachers reported that observation would be based on evidence, rather than administrator's subjective opinion. Indeed, instead of just marking a score on a checklist (as was the case with PDAS), principals were required to provide detailed scripting of the lesson and

attach specific pieces of evidence to their observation ratings on each of the components. As a result, teachers believed the process would be more "rigorous," "intense," and "structured." Administrators also reported that the new evaluation process helped decrease their own level of bias, "*PDAS had room for the individual observing you and I didn't agree with that. In INVEST, evidence has to be shown, which teachers like. It takes out any bias from what is observed… You focus on the facts. It's not about opinions"* (School 3, Principal).

Feedback/Growth. Given the increased observation requirements associated with INVEST, teachers generally anticipated receiving more detailed and frequent feedback on their performance. Unlike PDAS which was recorded manually, INVEST instituted a new online system, Teachscape, where principals could leave detailed feedback on teachers' performance aligned to specific components of the Danielson Framework. Despite the presence of these systems and structures, schools were overwhelmed by the timeline in the early months of implementation, which meant that many of teachers I interviewed had yet to receive an observation. As such, their perceptions of the feedback process remained primarily hypothetical in nature.

Principals and teachers both shared that the most significant benefit of INVEST would be its potential to increase dialogue about teaching practice. One teacher shared: *"I think that's really important for us as teachers to have that opportunity to tell them, you didn't see this but this is what I've been doing... I think it has opened up the communication lines, which is really positive"* (School 1, Teacher 1). Teachers reported several opportunities to share input, both during the pre-conference phase and through the

goal-setting and reflection processes. For many veteran teachers, this was the first time in years they had been asked to reflect on their performance. Some veterans found this process to be frustrating and time consuming, while others felt empowered by the opportunity to drive their own self-reflection. As one veteran teacher shares, "*I've never done this type of reflection before. It's good because it helped me actually stop and be honest with myself about where I need to improve"* (School 1, Teacher 6). I will explore this variation across individuals in greater detail in Chapter 5.

Phase 2: End of year

At the end of the year (in May), I interviewed the same subset of teachers and administrators in pilot schools to gather information on how teachers' perceptions had shifted over the course of the pilot year along the same themes identified in Phase 1 – level of understanding/purpose, system accuracy/fairness, and opinions on the quality of feedback/professional growth opportunities. This data was supplemented by the end of year survey data, which I used to compare pilot teachers' perceptions of INVEST to teachers remaining under the traditional PDAS system. As presented in Table 2-3, the survey collected information on teachers' perceptions of system design and implementation (outlined below). Some of the questions were asked of both pilot and non-pilot school teachers, while other questions were only asked of teachers in pilot schools. All measures were captured on a scale of 1-5 (1 being strongly disagree and 5 being strongly agree).

- ***Quality of Evaluation Measures (all teachers)*** – whether teachers agreed the evaluation measures were specific and clear, accurate and fair, comprehensive, and student-centered
- ***Fairness of Evaluation Process (all teachers)*** – whether teachers agreed the evaluation process was fair and accurately captured their performance
- ***Frequency of Evaluation (all teachers)*** – whether teachers agreed that evaluators spent adequate time observing them and meeting with them to discuss their practice.
- ***Number of Observations (all teachers)*** – the number of observations teachers reported receiving over the course of the year
- ***Number of Conversations (all teachers)*** – the number of conversations teachers reported receiving over the course of the year
- ***Quality of Growth and Feedback (all teachers)*** – whether teachers agreed that the evaluation system encouraged their professional growth, provided feedback that identified specific areas for improvement, and resulted in changes in practice
- ***Level of Understanding (pilot teachers only)*** – whether teachers agreed that the communication and training they received on INVEST helped to build their understanding of the new system
- ***Positive Goal-setting –*** whether teachers agreed that the goal-setting process helped them focus their efforts for the year and set more challenging goals

- ***Accuracy of INVEST Measures* –** whether teachers agreed that the Danielson Framework and Student Growth Percentiles measure were accurate and fair measures of their performance
- ***Positive Impact of INVEST* –** whether teachers agreed that INVEST provided specific feedback and support to improve teaching and would support teacher development

Since data was collected in May, I had expected that the system would have been fully implemented by this point of the year. However, I learned in interviews and informal conversations with district leadership that principals continued to struggle with implementation fidelity until the end of the year, and as such, had not always completed final end of year conversations by mid-May. As a result, though all teachers had more experience with the system than they did at the beginning of the year, some still had questions about how the system would play out for them at the end of the year.

Understanding/Purpose. Over the course of the year, district leadership attempted to respond to variation in teachers' initial perceptions of the system's purpose by offering additional INVEST training. In particular, they developed a series of online modules and an assessment on Student Growth Percentiles, which provided answers to many of the questions raised in the interviews and also created a series of presentations that administrators could use throughout the year with their teachers to build understanding of the system as a whole. Despite additional training, as demonstrated in Table 3-2, teachers' perceptions of the quality of ongoing communication throughout the year were slightly lower (M = 3.26) than they had been at the beginning of the year (M =

3.31), with only 51% of teachers reporting that the ongoing information they received about INVEST improved their understanding of the new system. In some cases, the additional information resulted in a better understanding of the rigor of the new system's expectations, which unintentionally heightened concern and frustration. This was particularly the case with SGPs, where after viewing the online modules, many teachers believed the student growth measure would not be able to control for factors outside of their control (e.g., student behavior, student attendance).

In spite of district leadership's efforts, teachers continued to have varying perceptions of the purpose of INVEST at the end of the year. As demonstrated in Table 3-2, at the end of the year, teachers were more likely to believe that INVEST would serve as an effective accountability tool (M = 3.41) than a tool for improving teaching (M = 3.09). For some teachers, this accountability was an important and necessary way to ensure improved student achievement, while for others, it was viewed as a tactic for demonizing teachers. One particularly frustrated teacher shared: *"INVEST has been used as a hammer to drive it all. INVEST is being used as a club against teachers, as a bullying tactic, as a weapon, so it's exacerbated problems that were already in existence"* (School 6, Teacher 2). Other teachers did not see accountability and improvement as mutually competing purposes: *"I guess the purpose of it was to pinpoint the needs in the classroom as far as the student growth and teacher growth. So they were trying to see whether or not your kids grew, not necessarily if they're perfect, but have they grown from year to year…and to support your growth as a teacher"* (School 3, Teacher 6). As demonstrated in Table 3-2, notwithstanding some teachers on either

extreme, close to half of surveyed teachers were neutral on whether INVEST would have an overall positive impact on the district (39%). Indeed, the modal category of teachers was fairly skeptical about the system's implementation and still in the process of forming their opinions.

Accuracy/ Fairness. As was the case at the beginning of the year, perceptions of the accuracy and fairness of the evaluation measures were central to teachers' overall attitudes toward INVEST. However, teachers' perceptions of the measures had changed over the course of the year. As demonstrated in Table 3-1, teachers in pilot schools had lower perceptions of both evaluation measures and processes (across all survey questions) compared to teachers in non-pilot schools. This result was somewhat surprising, given what many teachers and principals shared at the beginning of the year regarding the shortcomings of the prior PDAS evaluation system and the initial possibility of the new evaluation measures and processes under INVEST. Of particular significance, teachers in pilot schools rated the overall fairness of the new evaluation system at $M = 3.39$, compared to $M = 3.86$, for teachers in non-pilot schools, $p < .05$.

The interview data shed some light on what contributed to the shift in teachers' concerns over the accuracy and fairness of the new system. In general, teachers were still fairly positive about the specific domains of the Danielson measure. They maintained that the measure was "specific and evidence-based" and appreciated the "clarity of expectations" the rubric offered for evaluating their performance. However, after having received several observations (which had not yet happened at the beginning of the year), they expressed considerable frustration with Level 4 or the "Distinguished Level" of the

framework, sharing that the expectations were "unrealistic," "impossible to attain," and even "absolutely outrageous." The Distinguished Level required teachers to create student-centered classrooms, where students were responsible for taking ownership over their own learning process (through group and independent work, as well as student-driven questions). After realizing what these expectations meant in practice, many teachers did not believe they were reasonable for students who were often significantly below grade level.

Though teachers still had fairly positive perceptions of the Danielson measure overall (with the exception of Level 4 performance), they raised new concerns over the process of implementation, which contributed to overall perceptions of system fairness. One teacher shares, *"When I was observed, I didn't feel like everything that they saw reflected what I had to do in the classroom because depending on what day they walked in, I was doing different things. I don't feel like they got a very good picture of what I actually do in the classroom"* (School 4, Teacher 6). In particular, teachers (such as the one above) reported being concerned about the accuracy and usefulness of walkthroughs, which typically lasted for only 15 minutes. Even though these walkthroughs failed to capture a full lesson cycle, teachers were still scored on all components of the Framework. Additionally, INVEST considerably increased teachers' workload. At the end of the year, teachers were required to compile an artifact binder with detailed documentation of their performance on Domains 1 and 4. For many teachers, INVEST became synonymous with "increased paperwork" which they did not view as fair given already overwhelming demands on their time.

Feedback/Growth. At the beginning of the year, teachers in pilot schools had high hopes for the type of feedback and quality of support they would receive on the new system. However, due to challenges with the fidelity of implementation, many principals reported struggling to meet the new system requirements. At the end of the year, teachers in pilot schools rated the quality of feedback and opportunities for professional growth significantly lower than teachers in comparison schools. In particular, as demonstrated in Table 3-1, pilot teachers reported significantly lower perceptions of the feedback and opportunities for growth (M = 3.37) than comparison teachers (M = 3.64), $p < .001$. In pilot schools, two of the lowest scored survey items were the level of support offered by the new system (M = 3.01) and the system's ability to impact teacher development (M = 3.16).

As a result of implementation challenges, teachers did not typically receive the specific and actionable feedback they anticipated at the beginning of the year. Though teachers continued to believe that the Danielson Framework provided clear expectations, they did not generally report knowing how to effectively improve performance to meet the new and demanding standards (particularly Level 4 performance). Given their initially high expectations, many of the teachers I interviewed at the end of the year were frustrated that the system did not deliver on its promise of specific and actionable feedback. Despite these overall trends, the qualitative data suggest that there was considerable variation in implementation, which contributed to divergent results. In the section below, I will introduce some variation across teacher and school subgroups and revisit this in more detail in Part 2 of this dissertation.

Table 3-1

Teachers' Survey Perceptions of Evaluation in Pilot and Non-Pilot Schools

Measure	Overall Mean Scale (1-5)	Pilot Mean Scale (1-5)	Non-Pilot Mean Scale (1-5)
Quality of Evaluation Measures	3.77 (0.82)	3.53*** (0.88)	3.94*** (0.73)
Fairness of Evaluation Process	3.70 (0.91)	3.40*** (0.93)	3.91*** (0.83)
Frequency of Evaluation	3.83 (0.96)	3.68*** (0.99)	3.93*** (0.92)
Reported Number of Observations	4.16 (4.45)	3.91* (4.18)	4.34* (4.63)
Reported Number of Conversations	2.76 (2.68)	2.75 (1.87)	2.76 (3.14)
Quality of Feedback and Growth	3.54 (0.83)	3.38*** (0.88)	3.65*** (0.78)

Note. $N = 2662$. All survey questions were asked on a scale of 1-5 with 1 being Strongly Disagree and 5 being Strongly Agree. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

Table 3-2

Teachers' Survey Perceptions of INVEST-Specific Features in Pilot Schools

Measure	Overall Mean Scale (1-5)	% Strongly Disagree	% Disagree	% Neutral	% Agree	% Strongly Agree
Level of Understanding						
Initial understanding	3.31	6.78	16.64	23.61	44.52	8.46
At the beginning of the year	(1.06)					
Ongoing communication	3.26	6.48	17.04	25.65	45.19	5.65
Throughout the year	(0.98)					
Quality of observation training	3.33	6.11	12.21	30.34	45.05	6.29
Teachscape modules	(0.98)					
Teachscape online system	3.31	7.88	17.42	19.93	45.23	9.55
Ease of use	(1.11)					
Quality of SGP training	3.32	5.09	12.22	34.26	42.41	6.02
Student Growth modules	(0.98)					
Useful Goal-Setting						
Goal-setting focused efforts	3.31	5.46	14.81	29.17	44.26	6.30
	(0.98)					
Set challenging goals	3.01	8.62	22.24	33.83	29.84	5.47
	(1.04)					
Accuracy and Fairness of INVEST Measures						
Danielson Overall	3.06	9.06	15.65	38.45	33.21	3.63
	(1.00)					
Danielson Domain 1: *Planning and Preparation*	3.44	4.44	8.78	31.98	47.69	7.12
	(0.91)					
Danielson Domain 2	3.40	4.90	9.90	32.65	45.88	6.66
Classroom Environment	(0.93)					
Danielson Domain 3	3.32	5.37	11.75	33.95	42.92	6.01
Instruction	(0.95)					
Danielson Domain 4	3.41	4.54	10.19	31.88	46.15	7.23
Professional Responsibilities	(0.93)					
Student Growth Percentiles	2.93	9.65	21.80	38.78	25.14	4.64
	(1.02)					
INVEST Growth and Impact						
Quality of feedback	3.37	3.99	13.81	30.58	44.39	7.23
	(0.95)					
Level of positive support	3.01	7.98	20.50	38.78	27.83	4.92
	(1.00)					
Positive impact on my teaching	3.09	9.39	15.06	37.55	32.81	5.20
	(1.03)					
Positive impact on development	3.17	8.26	12.26	37.98	36.86	4.64
	(0.99)					
Positive impact on students	2.98	10.67	17.25	40.07	27.55	4.45
	(1.03)					
INVEST Overall Positive Impact	2.89	14.11	17.73	38.07	24.98	5.11
	(1.09)					

Note. $N = 1097$. All survey questions were asked on a scale of 1-5 with 1 being Strongly Disagree and 5 being Strongly Agree.

Section Two: Subgroup Analysis

Individual Variation

To be motivating, performance management systems must align with the expectancies and values of individual teachers. As a result, initial motivational responses to performance management policies will vary across subgroups and certain individuals will be more likely to improve practice over time. Both the qualitative and quantitative data suggest that teachers' perceptions of INVEST differed across dimensions of their effectiveness and experience. Teachers who did not reach the Highly Effective Level but felt their performance warranted that distinction were subsequently frustrated by the system. This was particularly the case for veteran teachers, who appeared not to be as open to the new system as novice teachers. This section explores this variation across subgroups of teachers.

Teacher Effectiveness. As demonstrated in Table 3-3 below, teachers who reached Level 4 (Highly Effective status) on the Danielson Framework tended to have better perceptions of the new evaluation system across the board than teachers at the lower levels of performance. In particular, Level 4 teachers viewed the evaluation measures as more accurate and likely to capture their teaching effectiveness, $M = 3.86$, when compared to Level 2 (Needs Improvement status) teachers, $M = 3.15$, $p < .001$. The contrast between Level 4 teachers and the other levels was even more pronounced for perceptions of the fairness of the evaluation process. Though on average, the mean perception of fairness of the new evaluation system was 3.41, Level 4 teachers were more

likely to believe the evaluation process was fair (M = 3.92) particularly compared to Level 1 teachers (M = 2.65) and Level 2 teachers (M = 2.74), $p < .001$. Interestingly, though Level 4 teachers were more likely to report that they received an adequate number of observations and conversations over the course of the year than teachers at other levels of performance, there were no statistically significant differences between the reported number of observations and conversations across levels of performance. Indeed, though the difference was not statistically significant, Level 1 teachers received more observations and conversations than their higher performing counterparts, suggesting that the issue was not observational frequency but rather, teachers' perceptions of observational accuracy. In terms of perceptions of the system's positive impact, Level 4 teachers were more likely to view INVEST as leading to opportunities for professional growth though these differences were not as pronounced as other system attitudes.

It is perhaps not surprising that teachers who reached higher levels of performance on INVEST were more likely to report that the system fairly captured their performance. Indeed, motivational theory would predict that we would value the accuracy of a system that affirms our personal competence. In interviews, the majority of teachers who had reached the Highly Effective status on the Danielson Framework shared that they felt validated for their hard work, which many believed had gone unrecognized under the prior PDAS evaluation system (since the majority of teachers received the highest ratings). In contrast, the veteran teachers who had always reached the highest level of performance under the PDAS system (Exceeds Expectations) but were not receiving Level 4 status on INVEST were more likely to be frustrated by the new system.

Table 3-3

Individual Variation in Survey Perceptions by Teacher Performance Level on Danielson Framework

Perceptions	Mean Scale (1-5)	Danielson Observation Rating			
		Level 1 *N*=17	Level 2 *N*=100	Level 3 *N*=806	Level 4 *N*=115
Teachers in All Schools	3.54	3.22	3.15	3.55	3.86
Quality of Measures***	(0.87)	(1.42)	(0.85)	(0.85)	(0.80)
Fairness of Process***	3.41	2.65	2.74	3.43	3.92
	(0.93)	(1.27)	(0.94)	(0.88)	(0.76)
Frequency of Evaluation***	3.70	3.09	3.38	3.72	3.97
	(0.98)	(1.29)	(1.10)	(0.95)	(0.90)
Reported Number of Observations	3.94	3.65	4.04	4.03	3.20
	(4.23)	(1.97)	(1.93)	(4.67)	(1.95)
Reported Number of Conversations	2.77	2.94	2.72	2.77	2.77
	(1.88)	(1.34)	(1.50)	(1.87)	(2.33)
Quality of Feedback and Growth	3.38	3.22	3.35	3.36	3.59
	(0.89)	(1.11)	(0.81)	(0.90)	(0.83)
Teachers in Pilot Schools	3.31	3.32	3.16	3.32	3.43
INVEST Level of Understanding	(0.82)	(0.95)	(0.80)	(0.82)	(0.80)
INVEST Positive Goal-Setting	3.17	3.47	3.06	3.17	3.26
	(0.91)	(1.07)	(0.93)	(0.91)	(0.90)
Accuracy of INVEST Measures***	3.27	3.50	3.01	3.27	3.45
	(0.79)	(0.98)	(0.76)	(0.79)	(0.70)
INVEST Growth and Impact*	3.20	3.54	3.04	3.20	3.31
	(0.82)	(0.87)	(0.84)	(0.82)	(0.75)
Positive Impact of INVEST*	2.89	3.41	2.73	2.87	3.08
	(1.09)	(1.06)	(1.08)	(1.09)	(1.09)

Note. Estimates were adjusted for multiple comparisons using the Scheffe method. The only differences that are statistically significant are between Level 4 and other levels of performance. *$p < .05$. **$p < .01$. ***$p < .001$

As the principal at the higher performing intermediate school described, *"we have winners in our building and we have people who are leaders and they all want to be distinguished, so that's the biggest thing that's been a challenge is hurt feelings"* (School 4, Principal). Rather than examine internal causes, many of these veterans attributed their lack of top performance to the unfairness of the system's measures and processes.

Though perceptions of evaluation varied considerably across teacher performance levels on the Danielson Framework, there were no statistically significant differences for any of the evaluation attitudes between teachers with different scores on the Student Growth Percentiles metric. In other words, while highly effective teachers on the Danielson Framework had more favorable attitudes towards the new evaluation system, highly effective teachers on the SGP metric did not react similarly. This can likely be attributed to the fact that teachers had yet to receive their SGP scores when they took the survey, so they were unaware of their performance on the metric. At the beginning of the year, Highly Effective teachers on the Danielson Framework did not appear to have more positive perceptions of the observation measure than their lower-performing counterparts. Rather, it was their actual success on the observation framework that appeared to influence their positive perceptions. If this logic holds, we would expect that once teachers see their SGP scores, those that reached Highly Effective status will have more positive perceptions of the accuracy of this measure as well.

Teacher Experience. Consistent with prior research (Johnson, 2005), first year teachers tended to have better perceptions of the new evaluation system's ability to help them grow their practice. Most notably, as demonstrated in Table 3-4, first year teachers

reported receiving more specific and quality feedback than teachers with additional years of experience ($M = 3.82$ compared to $M = 3.51$, $p < .001$), which contributed to the fact that they viewed INVEST as supporting their growth and development. This was perhaps not surprising given the requirements of the new system. Since first year teachers were on Track 1, principals were expected to observe and meet with them more frequently over the course of the year, and in practice, first year teachers reported receiving more observations (on average 4.69 compared to 4.12) and conversations (3.16 compared to 2.71) than their more experienced counterparts.

However, first year teachers' generally positive receptivity was not merely due to the fact that they received additional feedback on their practice under INVEST. Rather, they had a very different attitude towards the new system all together. As one first year teacher put it best, "*as first year teachers, we don't know any different than INVEST and we just want to be better*" (School 3, Teacher 4). Indeed, at the beginning of the year, first year teachers were very open to the new policy, because INVEST was the only system they had experienced and given their newness to the profession, they recognized the need to improve their performance. Principals, such as the one from School 5 quoted below, wished all their teachers would have reacted to INVEST in similar fashion to their novices:

> So I wish I had a building full of new teachers. Because they just eat it up. They want to be better. They want to know. They want to make sure every i is dotted and every t is crossed and they're fresh and energetic and they just want to know what they have to do to do it right. Those are the ones that are asking all the questions because they just want to know what do I need to do to be better because I know I have a lot to learn. And this system really teaches them. PDAS just wasn't that kind of system. It wasn't laid out that way.

As this quote demonstrates, first year teachers' initially positive mindsets were reinforced by the additional feedback they received under the new system.

Table 3-4

Individual Variation in Survey Perceptions by First Year Teacher Status

Evaluation Attitudes	Mean Scale (1-5)	Experience	
		First Year *N*=183	2+ Years *N*=2284
Teachers in All Schools			
Quality of Evaluation Measures*	3.77 (0.82)	3.88* (0.79)	3.76* (0.82)
Fairness of Evaluation Process	3.70 (0.91)	3.80 (0.89)	3.69 (0.91)
Frequency of Evaluation	3.83 (0.96)	3.86 (1.00)	3.83 (0.95)
Number of observations*	4.16 (4.52)	4.69 (3.05)	4.12 (4.62)
Number of conversations*	2.76 (2.60)	3.16 (1.87)	2.72 (2.75)
Quality of Feedback and Growth***	3.53 (0.83)	3.82 (0.78)	3.51 (0.83)
Teachers in Pilot Schools		N=81	N=935
INVEST Level of Understanding	3.29 (0.82)	3.26 (0.84)	3.29 (0.82)
INVEST Positive Goal-Setting	3.15 (0.91)	3.28 (0.86)	3.14 (0.92)
Accuracy of INVEST Measures	3.25 (0.79)	3.39 (0.70)	3.24 (0.80)
INVEST Growth and Impact*	3.19 (0.82)	3.39 (0.82)	3.17 (0.81)
Positive Impact of INVEST	2.88 (1.09)	3.06 (1.13)	2.86 (1.09)

Note. Estimates were adjusted for multiple comparisons using the Scheffe method. *$p < .05$. **$p < .01$. ***$p < .001$. The statistically significant differences are between first year teachers and their more experienced counterparts.

School Variation

In addition to variation at the individual level, research has also demonstrated that teachers' responses to new systems can be influenced by school context. Though certain individuals may react differently within the same school, in the aggregate, teachers' responses will likely vary depending on the type and performance level of the school. The quantitative and qualitative data suggest that teachers' perceptions of INVEST differed across level of schooling, and to a lesser extent, by school performance. This section explores this variation across subgroups of schools.

School Level. Both sources of data suggest that teachers at the high school level (both ninth grade and senior high school) had lower perceptions of INVEST than other levels of schooling. Ninth grade teachers reported receiving fewer observations and conversations than teachers in lower levels of schooling, which confirms qualitative data that ninth grade principals had more significant challenges with implementation fidelity. Both ninth grade principals I interviewed shared that they had struggled to maintain the implementation timeline due to their many other responsibilities. Based on interview data, it appeared that principals at higher levels of schooling had extra responsibilities when compared to their counterparts at elementary schools; however, it is not clear what led to these differing expectations across school levels.

High school teachers also appeared to react differently to the new system expectations regardless of the frequency of their observation. As demonstrated in Table 3-5, high school teachers reported lower perceptions of understanding of INVEST, less investment in goal-setting under the new system, and more concerns over the quality of

Table 3-5

Variation in Teachers' Survey Perceptions by School Level

Evaluation Attitudes	Mean Scale (1-5)	Level					
		Pre-K	Elem	Inter-mediate	Middle	Ninth	High School
Teachers in All Schools		*N*=183	*N*=967	*N*=355	*N*=369	*N*=139	*N*=523
Quality of Evaluation Measures	3.77 (0.82)	3.75 (0.81)	3.81 (0.85)	3.83 (0.73)	3.71 (0.84)	3.72 (0.76)	3.73 (0.81)
Fairness of Evaluation Process*	3.70 (0.91)	3.76 (0.85)	3.76 (0.92)	3.64 (0.88)	3.65 (0.93)	3.47 (0.82)	3.69 (0.94)
Frequency of Evaluation***	3.83 (0.96)	3.93 (0.82)	3.92 (0.90)	3.83 (0.95)	3.75 (0.98)	3.54 (1.04)	3.76 (1.03)
Number of Observations	4.16 (4.46)	3.75 (1.88)	4.36 (5.40)	4.29 (2.52)	4.05 (4.96)	3.18 (2.17)	4.33 (4.43)
Number of conversations	2.76 (2.69)	2.55 (1.95)	2.69 (1.99)	2.96 (1.87)	2.78 (2.60)	2.31 (1.re)	3.00 (4.40)
Quality of Feedback and Growth***	3.54 (0.84)	3.52 (0.80)	3.61 (0.83)	3.65 (0.80)	3.45 (0.82)	3.29 (0.88)	3.49 (0.86)
Teachers in Pilot Schools		*N*=81	*N*=362	*N*=251	*N*=154	*N*=117	*N*=82
INVEST Level of Under-Standing***	3.31 (0.82)	3.22 (0.74)	3.33 (0.84)	3.48 (0.75)	3.31 (0.79)	3.24 (0.84)	2.93 (0.82)
INVEST Positive Goal-Setting***	3.16 (0.91)	3.26 (0.80)	3.14 (0.97)	3.35 (0.82)	3.13 (0.86)	3.14 (0.90)	2.73 (0.90)
Accuracy of INVEST Measures	3.26 (0.79)	3.27 (0.91)	3.21 (0.80)	3.43 (0.71)	3.26 (0.78)	3.27 (0.77)	2.98 (0.82)
INVEST Growth and Impact***	3.20 (0.81)	3.17 (0.77)	3.18 (0.83)	3.42 (0.72)	3.18 (0.86)	3.18 (0.74)	2.79 (0.88)
Positive Impact of INVEST***	2.89 (1.09)	2.90 (1.09)	2.78 (1.08)	3.23 (1.02)	2.75 (1.16)	2.97 (0.91)	2.54 (1.15)

Note. Estimates were adjusted for multiple comparisons using the Scheffe method. *$p < .05$. **$p < .01$. ***$p < .001$. The only differences that are statistically significant are between Ninth Grade and High School and other levels of schooling.

feedback they received during the evaluation process. As a result, it is perhaps not surprising that high school teachers were significantly less likely to view the new system

as supporting professional growth, $M = 2.79$, and less likely to have a positive impact on the Aldine ISD, $M = 2.54$.

Though there was variation, the high school teachers I interviewed tended to be more skeptical about INVEST's usefulness and questioned its potential to have a positive impact on student learning. One skeptical high school teacher shared:

> I think initially for myself I thought, wow, this would be really good in the elementary setting. And then for it to grow as they grow in the system because I have high school students now that are juniors, they would be like, what? They have not had that environment of working together and taking the ownership. I'm sure there's a way to rein it back in, but for them, especially if you have high school students that are on the fence about their education, they would be really hesitant and that will become another barrier and then we're talking about evaluating the teacher and the students' reluctance would be a great factor for me. Definitely with the elementary kids and then being ground level and their little natures anyway is to want to work together (School 6, Teacher 4).

As this quotation demonstrates, high school teachers' concerns were often rooted in their belief that high school classrooms should be structured differently than elementary classrooms, given the age and needs of the students. Indeed, high school teachers were more likely to report concerns over student motivation, which contributed to their concerns over the feasibility of creating student-led classrooms.

School Performance. Based on the state of Texas's rating system, Aldine schools received one of three designations at the end of the 2011-2012 school year – Acceptable (average performance compared to other schools in the state), Recognized (above average performance compared to other schools in the state), and Exemplary (exceptional performance compared to other schools in the state). Both the quantitative and qualitative data suggested that teachers at higher performing schools appeared to have lower

Table 3-6

School Variation in Survey Perceptions by School Performance Rating

Evaluation Attitudes	Mean Scale (1-5)	School Performance Rating		
		Acceptable	Recognized	Exemplary
Teachers in All Schools		*N*=481	*N*=1134	*N*=252
Quality of Evaluation Measures	3.78 (0.81)	3.80 (0.77)	3.74 (0.83)	3.89 (0.86)
Fairness of Evaluation Process**	3.71 (0.91)	3.75 (0.88)	3.63 (0.92)	3.87 (0.95)
Frequency of Evaluation**	3.83 (0.96)	3.80 (1.01)	3.81 (0.91)	4.03 (0.91)
Quality of Feedback and Growth	3.55 (0.83)	3.53 (0.82)	3.53 (0.83)	3.69 (0.87)
Teachers in Pilot Schools		*N*=207	*N*=701	*N*=66
INVEST Level of Understanding	3.33 (0.82)	3.29 (0.86)	3.36 (0.82)	3.13 (0.72)
INVEST Positive Goal-Setting**	3.17 (0.91)	3.13 (0.91)	3.22 (0.89)	2.85 (1.08)
Accuracy of INVEST Measures*	3.28 (0.77)	3.24 (0.79)	3.30 (0.76)	3.07 (0.79)
INVEST Growth and Impact**	3.22 (0.81)	3.16 (0.85)	3.27 (0.78)	2.92 (0.84)
Positive Impact of INVEST**	2.91 (1.08)	2.87 (1.12)	2.96 (1.06)	2.52 (1.07)

Note. Estimates were adjusted for multiple comparisons using the Scheffe method. Teachers in Pre-K centers and one new school are excluded from the analysis because they did not have performance data in the 2011-2012 school year.
**p<.05. **p<.01. ***p<.001*. The only differences that are statistically significant are between Recognized and Exemplary Schools.

perceptions of INVEST on certain measures. As demonstrated in Table 3-6, though there were no significant differences between Acceptable and Recognized schools, teachers in Recognized schools tended to have better perceptions of INVEST compared to teachers in Exemplary schools. In particular, teachers at Recognized schools reported being more

likely to engage in goal-setting (M = 3.22) compared to teachers at Exemplary schools (M = 2.85), $p < .01$ and believed that INVEST measures accurately captured performance, (M = 3.30 compared to M = 3.07), $p < .01$. As a result, teachers in Recognized schools had better overall perceptions of INVEST's potential for growth and impact than teachers in Exemplary schools.

At first glance, this finding is somewhat puzzling. Indeed, policymakers would probably expect higher performing schools to have more positive reactions to new systems. However, upon further investigation, it is consistent with what the principal in School 4 shared about the culture on higher performing campuses. Recall her statement that *"we have winners in our building and we have people who are leaders and they all want to be distinguished, so that's the biggest thing that's been a challenge is hurt feelings."* Based on interview data, it appeared that in previous years, higher-performing schools had more teachers rated at the top level of the PDAS system. During the pilot year of implementation, teachers at the high-performing schools I visited remarked that the principal had very high expectations for performance. As a result, high performing pilot schools seemed to have more teachers who did not reach their desired level of performance (Level 4) under the new system, which likely contributed to less positive perceptions.

Summary

In sum, teachers' attitudes towards the new system at the beginning of the year were mixed and subsequently shifted over the course of the year. Across the board, teachers in the early months of implementation were overwhelmed by the timeline and

increased expectations associated with INVEST. Yet despite the additional workload, the majority of teachers appreciated the clarity and comprehensive nature of the new observation measure (the Danielson Framework) and the detailed and evidence-driven nature of the new observation process (utilizing the online Teachscape system). Seemingly, their attitudes toward the new system were influenced by their understanding of the system's purpose. Teachers who believed INVEST was designed to support their professional growth were more likely to report being invested in the system's potential, compared to teachers who believed the system was designed primarily as an accountability tool. Due to challenges with the fidelity of implementation in the first few months of the year, teachers' attitudes were still primarily based on their initial understanding and yet to be influenced by substantial experience with the new system.

At the end of the year, teachers in pilot schools had lower perceptions of evaluation when compared to teachers in non-pilot schools across all survey metrics – quality of evaluation measures, fairness of the evaluation process, frequency of evaluation, and quality of feedback and growth. One plausible explanation for this finding could be that the substance of INVEST actually fell short of the quality of the previous PDAS system. Though some of the teachers I interviewed were unquestionably frustrated by aspects of the new system (e.g., unrealistic performance expectations, increased workload), only two of the 36 interviewed teachers shared that they preferred the PDAS system over INVEST, making this an unlikely possibility. However, certain design features of INVEST (i.e., level of understanding of the purpose, perceived

accuracy of the new evaluation measures, and the quality of feedback and support) did indeed appear to influence teachers' responses to the new policy.

Another plausible explanation could be related to teachers' individual responses to change. The modal survey response for the majority of teachers' perceptions of INVEST was "neutral," suggesting that many teachers were skeptical and still unsure of the potential impact of the new system. As with any significant change, we might naturally expect initial resistance. However, there was considerable variation in how teachers responded to INVEST. In particular, subgroup analysis revealed that highly effective and first year teachers appeared more likely to respond positively to the new system's expectations.

A final explanation is rooted in the implementation process at the school level. At the beginning of the year, many teachers and principals had high expectations for the system's promise; however, INVEST did not live up to its potential for many teachers due to challenges with the fidelity of implementation across schools. As both the quantitative and qualitative data made clear, high school teachers were less likely to report being satisfied with the new evaluation system and were particularly concerned about the attainability of the INVEST measures at the secondary level. The pattern of perceptions based on school performance is less clear, but evidence does seem to suggest that teachers at higher performing schools may be more frustrated by failing to meet the expectations of the new system and in turn, had less positive attitudes towards INVEST.

In short, though there were trends in overall teacher attitudes, there was considerable variation at the individual, school, and system level that appeared to

contribute to teachers' responses to the new system. I will explore how these three sources of variation influenced outcomes in Part Two of this dissertation, but first, in Chapter 4, I turn my attention to how initial attitudes toward INVEST translated into teacher motivation, effectiveness, and retention.

CHAPTER 4: OVERALL SYSTEM IMPACT

The descriptive analysis presented in Chapter 3 offers a foundation for understanding the implementation of the new system. It makes clear that teachers' perceptions of INVEST were influenced by their beliefs in its purpose, their judgments of the accuracy and fairness of the evaluation measures and process, and the quality of feedback and opportunities for professional growth. When teachers presumed that INVEST was designed to enhance teaching and learning, as opposed to serving primarily as an accountability tool, they were initially more receptive to the new expectations. Regardless of perceptions of the system's purpose at the beginning of the year, teachers generally appreciated the Danielson Framework's clear and specific expectations for performance and anticipated receiving more detailed feedback and engaging in additional dialogue about their practice.

However, on average, at the end of the year, pilot teachers had lower perceptions of the evaluation system on the majority of metrics assessed on the end of year survey. Indeed, after experiencing system implementation, many teachers (particularly veterans who were no longer performing at the top of the evaluation system) conveyed frustration with the unattainability and unfairness of INVEST's requirements. Though there was considerable variation across individual profiles of teachers and school contexts (which will be discussed in Part Two), these descriptive results suggest that the new system would not have an overall positive impact on teacher outcomes – i.e., motivation, effectiveness, and retention.

This chapter builds on these descriptive results by examining the impact of the new INVEST system on teacher motivation, effectiveness, and retention. Using difference-in-differences (DID) analysis, I estimated the treatment effect of the new system by comparing pilot schools after the treatment (INVEST) both to the pilot schools before treatment and to the non-pilot schools in the district. As discussed in Chapter 2, DID presumes that in the absence of INVEST, the pilot schools would have continued on the same trajectory as the non-pilot schools. Though pilot and non-pilot schools were equivalent at baseline on key observables, these impact estimates have the potential to be biased because I cannot account for possible time-varying unobservable characteristics between the groups of schools (such as features of the school climate). To attempt to account for differences in the initial composition, I controlled for individual-level characteristics (i.e., gender, ethnicity, certification type, and years of experience) in the models. Since the data were collected at the individual level but the intervention was at the school-level, I clustered my standard errors at the school level and conducted my analysis with and without school fixed effects, and present both sets of results. Though this chapter relies primarily on quantitative data, I supplement the discussion of each impact – motivation (Section One), effectiveness (Section Two), and retention (Section Three) – with overall trends from the qualitative interview data.

Section One: Motivation

This section explores the impact the pilot had on teachers' expectancy and value over the course of the school year. As discussed in Chapter 1, expectancy-value theory links motivational choices to two sets of beliefs: an individuals' expectation of success

(expectancy) and the importance or value the individual associates with specific actions (value). At its most basic level, individuals ask two central questions when determining initial motivation: "Can I do the task?" and "Do I want to do the task?" If teachers doubt whether they are able to reach desired expectations, they will be unlikely to change behavior. Further, teachers who believe in their ability to reach expectancy but do not value the action itself or outcomes associated with it are also unlikely to be motivated to change behavior.

Expectancy (Can I do the task?). Expectancy theorists distinguish between two general expectancies – self-efficacy and outcome expectations. Self-efficacy captures an individual's (e.g., teacher's) belief about their own level of competence (e.g., personal expectancy) while outcome expectations document whether an individual thinks a given behavior will lead to certain outcomes (e.g., system expectancy). In the case of teacher evaluation, teachers would have high outcome expectancy if they believed their improvements would be recognized by the system and high overall expectancy if they believed in their ability to have a significant impact on their students (regardless of the new system).

Value (Do I want to do the task?). Motivational theorists contend that the perceived value of any given activity can be determined by four constructs – (1) the intrinsic interest (or enjoyment) one expects to get from a specific task, (2) attainment value, or the extent to which a task is consistent with an individual's self-image, (3) the utility value of the task for achieving long-range goals, and (4) the perceived cost of a particular action. The first construct, intrinsic interest, is influenced by a desire to be

autonomous, competent, and connect with others and in the case of this analysis, is most in line with teachers' personal value for their work. The second, third, and fourth constructs are more related to an extrinsic desire to be recognized or achieve some greater end – in the case of this analysis, similar to system value.

Quantitative Results

Descriptive Evidence

Table 4-1 captures the unadjusted trends in motivation (expectancy and value respectively), for all Aldine ISD schools, the pilot schools, and the non-pilot schools. Prior to the implementation of INVEST (in fall 2012), levels of personal expectancy and personal value in pilot schools were virtually identical to non-pilot schools. After the implementation of the new system, pilot schools' level of expectancy appeared to drop slightly (-0.04), compared to non-pilot schools where level of expectancy slightly increased (0.02). More notably, teachers' level of expectancy on INVEST (system expectancy) was considerably lower than any of the other measures (M = 3.24, SD = 1.14), which may mean that lower system expectancy is contributing to lower personal expectancy in pilot schools. Unlike expectancy, teachers' value in pilot schools remained consistent over the course of the year, while value in the non-pilot school slightly increased (0.02). Pilot teachers' level of value for performing well on the system (system value) was similar to their personal value at both the beginning and end of the year (M = 4.21, SD = 0.83).

Table 4-1

Descriptive Statistics: Teacher Self-Reported Motivation (Captured From Survey Data)

Measure (On a Scale of 1-5)	Pilot	Non-Pilot
Personal Expectancy (start of year)	3.98	3.98
(Belief in Ability)	(0.58)	(0.56)
Personal Expectancy (end of year)	3.94	4.00
	(0.60)	(0.58)
Personal Value (start of year)	4.21	4.20
(Value for the Work)	(0.61)	(0.60)
Personal Value (end of year)	4.21	4.22
	(0.58)	(0.58)
System Expectancy (end of year)	3.24	--
(Belief in Ability on INVEST)	(1.14)	
System Value (end of year)	4.21	--
(Value for INVEST)	(0.83)	

Note. N = 2662 with the exception of System Expectancy and System Value where N = 1097.

Impact Estimates

Using the difference-in-differences identification strategy, I estimated the effect of the evaluation pilot on teachers' personal motivation, measured as two separate constructs of expectancy and value. For each measure, I ran four models. The first two models did not include school effects, meaning they did not account for school-level characteristics that may contribute to observed outcomes. The second two models included school fixed effects to adjust for these characteristics and created a more robust estimate of the pilot's impact on motivation. For each of these sets of models, I ran the analysis with and without individual-level covariates (i.e., race, gender, certification, years of schooling) to assess whether the impact of the pilot was significant after controlling for individual level variation.

As illustrated in Tables 4-2 and 4-3, the impact of the pilot was inconsistent across the two sources of personal motivation – expectancy and value. For personal expectancy, summarized in Table 4-2, the relative change in pilot teachers' expectancy at the end of the year (the "difference in differences") was -0.11 standard deviation units (or 0.06 in unstandardized units) and was statistically significant ($p < .05$) without including school-level effects or individual-level controls. In other words, at the end of the year, teachers in pilot schools had lower personal expectancy (or belief in their ability to impact students) as a result of having participated in the pilot. As demonstrated in Model 3, this estimate remained robust even with the inclusion of school-level fixed effects. However, when controlling for individual-level characteristics, as done in both Models 2 and 4, the impact estimate was no longer significant at the $p < .05$ level, but remained significant at the $p < .10$ level. Table 4-3 demonstrates that the pilot also had a negative (though smaller) impact on teachers' personal value (importance they place on their work) and that the relative changes in pilot teachers' value at the end of the year (the "difference in differences") was -0.04 standard deviation units (or 0.02 in unstandardized units). However, this impact was not statistically significant across any of the four models. Notably, as was the case with expectancy, these estimates did not appear to be influenced by the inclusion of school-level effects.

Analyses of both personal teacher expectancy and value revealed that the change in motivation differed across teachers with various characteristics. White teachers had lower growth in personal expectancy than their non-white colleagues resulting in end of year scores of $M = 3.87$, $SD = .61$ and $M = 4.03$, $SD = .57$ respectively and placed less

value on their teaching over the course of the year, with end of year scores of $M = 4.18$, $SD = .59$ compared to $M = 4.24$, $SD = .57$. Traditionally certified teachers increased their personal value for their work of the course of the year more significantly than their alternatively certified counterparts, $B = 0.12$ (.03). These differences are relatively small in magnitude but the impacts were statistically significant, $p < .001$. Since performance data on the Danielson Framework only existed for pilot schools, it could not be included in the analysis. However, as suggested in Chapter 3, teachers' performance on the new system appeared to influence their personal motivation to improve performance. I will explore this individual-level variation in more detail in Chapter 5.

Table 4-2

Pilot's Impact on Teachers' Self-Reported Personal Expectancy

	Without School Effects		With School Effects	
Measure	Model 1 *Impact of the Pilot*	Model 2 *With Individual Controls*	Model 3 *Impact of the Pilot*	Model 4 *With Individual Controls*
Impact of INVEST (Pilot*Year)	-0.10 (0.05)*	-0.09 (0.05)	-0.11(0.06)*	-0.09(0.06)
Pilot	0.00 (0.04)	-0.02 (0.04)	-0.20(0.12)	-0.32(0.27)
Year	0.02 (0.03)	0.01 (0.03)	0.02(0.03)	0.01(0.03)
Gender *(female)*		0.05 (0.03)		0.00(0.03)
Ethnicity *(white)*		-0.23 (.03)***		-0.23(0.04)***
Certification *(traditional)*		-0.01 (0.03)		-0.03(0.04)
Years of Experience		0.00 (0.01)		0.02(0.02)

Note. N = 2662. All continuous variables have been standardized.
**p < .05. **p < .01. ***p < .001*

Table 4-3

Pilot's Impact on Teachers' Self-Reported Personal Value

Measure	Without School Effects		With School Effects	
	Model 1 *Impact of the Pilot*	Model 2 *With Individual Controls*	Model 3 *Impact of the Pilot*	Model 4 *With Individual Controls*
Impact of INVEST (Pilot*Year)	-0.04(0.05)	-0.02(0.05)	-0.06(0.05)	-0.03(0.05)
Pilot	0.01(0.05)	-0.02(0.04)	0.28(0.19)	0.71(0.41)
Year	0.04(0.03)	0.03(0.03)	0.05(0.03)	0.03(0.03)
Gender *(female)*		0.04(0.04)		0.00(0.03)
Ethnicity *(white)*		-0.13(0.03)**		-0.12(0.03)***
Certification *(traditional)*		0.12(0.03)**		0.10(0.03)***
Years of Experience		-0.02(0.02)		-0.01(0.02)

Note. $N = 2662$. All continuous variables have been standardized.
*$p < .05$. **$p < .01$. ***$p < .001$

Personal motivation for teaching was related to, but not synonymous with, motivation to perform well on the new evaluation system. As noted above, teachers' self-efficacy or belief in their own level of competence (e.g., personal expectancy) is distinct from whether they believe a given behavior will lead to certain outcomes (e.g., system expectancy). As demonstrated in Table 4-4, teachers' personal expectancy and system expectancy are significantly correlated but the magnitude is relatively small in size, $r = .23$, $p < .001$. This suggests that teachers' belief in their own ability to impact students is only moderately associated with whether they believe they can perform well on the new evaluation system itself. The same trends apply to personal and system value. While teachers may value being good teachers more likely for intrinsic reasons (personal value), this does not necessarily translate into a desire to be recognized by the new evaluation

system (system value); indeed, the correlation between the two measures is significant but of relatively small magnitude, $r = .29$, $p < .001$. Interestingly, teachers' personal expectancy and value have a higher correlation, $r = .45$, $p < .001$, than their system expectancy and value, $r = .26$, $p < .001$. These trends will be explored in further detail in the qualitative section below.

Table 4-4

Correlations between Teachers' Personal Motivation and System Motivation

Measure	1.	2.	3.	4.
1. Personal Expectancy	--	0.45***	0.24***	0.26***
2. Personal Value		--	0.23***	0.29***
3. System Expectancy			--	0.26***
4. System Value				--

Qualitative Results

Teachers' responses to the new system varied considerably both across and within schools, which will be discussed in greater depth in Part 2. Despite this variation, there are several trends in teachers' personal motivation that help elucidate the quantitative analysis described above.

Expectancy (Can I Do It?)

As discussed in Chapter 3, teachers reported that they appreciated the Danielson Framework's comprehensive nature and specifically outlined expectations. By providing a "clear roadmap," teachers knew what was expected of their performance, which led to relatively high levels of expectancy at the beginning of the year. As one teacher shared, *"That is like my Bible, you know? My direction. Without this rubric, you don't know what*

to do. You don't know your expectation" (School 6, Teacher 2). However, after the system was implemented, teachers began to view the expectations as unachievable, which lowered system expectancy. Given the low percentage of teachers reaching Level 4 status (only 14% in the district as a whole and as low as 0% in some schools), their perceptions of the challenge associated with reaching Level 4 behavior were quite rational. According to one teacher, the principal essentially told teachers not to bother trying, *"maybe it was the way it was presented to us by our administration, but they made it seem like it was going to be almost impossible to get a 4"* (School 5, Teacher 3). This lowered system expectancy also affected teachers' personal expectancy as educators by making them feel inadequate. In part, this was due to the terminology associated with the new system, which referred to teachers meeting standards as "effective" or "proficient," which several teachers found to be particularly demeaning: *"because the word proficient, even when I was a first year teacher, kind of sounds like I'm just average. I'm barely good enough. That's discouraging"* (School 6, Teacher 1).

Value (Do I Want To Do It?)

Despite these frustrations, the majority of teachers expressed a strong desire to reach the top level of the evaluation system (Level 4 performance). For many veteran teachers, this was a matter of pride; indeed, achieving highly effective status was necessary to maintain their self-image. One seasoned veteran compared his system value to his work with students, *"We model that in the classroom, we want our students to be the best, and we provide them opportunities in the classroom to be the best and to get A's and be those distinguished scholars. It should be no different with appraisals"* (School 5,

Teacher 1). Teachers' frustration with the new system often resulted from not meeting expectations, which indicated a high level of initial extrinsic attainment value. One veteran teacher who did not reach Level 4 performance shared, *"I thought I was doing everything that was highly effective. I didn't change anything I've done in the last year when I was always such a strong teacher. I've always been exceeds... It's very discouraging"* (School 2, Teacher 1).

Though teachers generally valued performing well on the system, this did not necessarily result in changes in the personal value they placed on their teaching and, in some cases, may have contributed to lower personal value. Many teachers reported that they were already motivated to improve their teaching and that evaluation would not have much of an impact on their personal value for teaching in either direction: *"I don't really think an evaluation should be a motivation to be a good teacher. I think that's just part of my job"* (School 5, Teacher 4). Though teachers may have continued to value teaching regardless of the new system, INVEST often led to a significantly intensified workload, which decreased many teachers' level of enjoyment with their daily experience as educators. As discussed in Chapter 3, these requirements included watching 16 hours of video (in a prescribed timeframe) and submitting specific documents to provide evidence of mastery on the Framework. When I asked teachers to describe INVEST, "more paperwork" was a fairly common initial response. One teacher described how the increased workload limited her ability to focus on what mattered most: teaching; *"I don't know. I'm going to be honest here. I just think it's becoming a bit much. I think what we're doing is we're getting away from what the root cause is, and I think it's learning…*

INVEST means I do more paperwork than teaching" (School 3, Teacher 6). Teachers generally did not believe that the extra work (or cost) aligned with valued outcomes, which for some, decreased their level of interest in their work: *"it's extra responsibilities being held over my head and keeps me from focusing on my teaching. It's been a huge source of stress"* (School 5, Teacher 2).

Section Two: Effectiveness

Given INVEST's limited (and slightly negative) effect on teachers' expectancy, we would likely not expect to see positive changes in teachers' effectiveness as a result of the new system. However, as theory indicates, teachers' initial motivation is distinct from the volition necessary to improve performance and sustain changes over time. Unfortunately, I do not have a strong measure of teachers' volition, so as a proxy I will use the end of year survey question which asked teachers to indicate whether they had implemented changes in their practice as a result of the new evaluation system. To assess effectiveness, teachers were evaluated on observations over the course of the year using the Danielson Framework for Teaching and on their impact on student growth using Student Growth Percentiles.

Danielson Framework (Danielson). The Framework for Teaching consists of 22 components divided into 4 broader domains – Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities. Over the course of the year, teachers received several walkthroughs depending on their years of experience and performance (novices and ineffective teachers received additional walkthroughs) and at least one formal observation. At the end of the year, these scores were weighted (50%

walkthroughs, 50% formal observation) to calculate a final score on each of the four domains and then an average score (on a scale of 1-4). Teachers were also given an overall rating that differentiated four levels of performance – Ineffective (Level 1), Needs Improvement (Level 2), Effective (Level 3), and Highly Effective (Level 4). Since this measure was only available in pilot schools at the end of the year, it cannot unfortunately be used as a measure in the impact analysis.

Student Growth Percentiles (SGPs). To measure teacher performance based on student growth, the district used a student growth percentile measure based on the Colorado Growth Model. The model compared the change in each student's achievement score to all other students in Aldine who had similar achievement scores in the previous year and then assigned the teacher an overall SGP score (on a scale of 1-100) based on the median SGP of their students. Because the measure could only be calculated for teachers in tested grades and subjects, the sample of teachers for which the measure was available is quite limited. Educators outside of tested subjects set Student Growth Objectives (SGOs) to measure their students' progress over the course of the year; however, given challenges with implementation, these results were not available. Given the limited sample for which SGPs are available (teachers in grades 3 to 9 in tested subjects), analysis of this data was not restricted to teachers who completed both surveys.

At the end of the year, the "Final INVEST Rating" was supposed to be drawn from scores on both observation and student growth (either student growth percentiles or student growth objectives). To be *Highly Effective or Effective* overall, teachers would need to be rated *Highly Effective or Effective* (respectively) on both measures, and

teachers would be scored as *Needs Improvement* or *Ineffective* if they received this rating in either of the measures. However, since the district leadership decided to only use the Danielson Framework for consequence (i.e., to put teachers on a professional growth plan) in the pilot year, many principals did not share SGP data with their teachers. It is not possible to ascertain how many teachers viewed their SGP results, but according to the data consultant, very few teachers had logged into the system and of the 36 teachers I interviewed, only two had seen their results. Given the limited sample and lack of use (as well as some methodological challenges which will be discussed in Chapter 6), the SGP effectiveness measure has some notable limitations as a measure of pilot impact; however, it is the only teacher effectiveness measure that is available prior to and post INVEST implementation.

Quantitative Results

Descriptive Evidence

Table 4-5 summarizes the effectiveness results (SGPs and Danielson respectively), for all Aldine ISD schools, the pilot schools, and the non-pilot schools. Prior to the implementation of INVEST (in the 2011-2012 school year), teachers' SGP ratings in pilot schools (M = 50.96, SD = 13.04) were slightly higher than non-pilot schools (M = 49.70, SD = 12.25). After the implementation of the new system, pilot teachers SGP ratings dropped slightly (-0.07), compared to non-pilot schools where teachers' SGP scores slightly increased (+0.95). In pilot schools, teachers' scores on the Danielson Framework averaged 3.19 with a fairly restricted range (as most teachers scores hovered around an average of 3). As far as reported changes in practice, teachers

in non-pilot schools had higher scores (M = 3.42, SD = .93), compared to teachers in pilot schools (M = 3.27, SD = .99), $p < .05$, which may be one possible explanation for why they saw slightly improved effectiveness scores.

Table 4-5

Descriptive Statistics: Teacher Effectiveness and Reported Change in Practice

Measure	Pilot	Non-Pilot
Danielson Effectiveness (end of year) *(On a scale of 1-4)*	3.20 (0.37)	--
SGP Effectiveness (previous year) *(On a scale of 1-100)*	50.96 (13.04)	49.70 (12.25)
SGP Effectiveness (end of year) *(On a scale of 1-100)*	50.89 (12.96)	50.65 (13.55)
Reported Changes in Practice *(On a scale of 1-5)*	3.27* (0.99)	3.42* (0.93)

Note. N = 906 for SGPs (end of year). N = 1097 for Danielson Effectiveness.
*$p < .05$. **$p < .01$. ***$p < .001$

Impact Estimates

As with motivation, I used difference-in-differences analysis to estimate the impact of the INVEST pilot on teachers' effectiveness, measured by Student Growth Percentiles, and ran four models to control for both school-level and individual-level effects. As demonstrated in Table 4-6, the pilot did not have a significant impact on teachers' effectiveness on Student Growth Percentiles. The relative change in pilot teachers' SGP effectiveness at the end of the year (the "difference in differences") was -0.08 standard deviation units (or -1.02 in unstandardized units) without including school-

level effects or individual-level controls. As the descriptive data suggested, teachers in pilot schools had lower effectiveness on SGPs at the end of the year (and less growth from the beginning of the year) than teachers in non-pilot schools. However, this "difference in differences" estimate was not statistically significant with or without school effects in the model. Additionally, the estimate of the pilot's impact became slightly positive (though not statistically significant) when controlling for individual-level characteristics in both Models 2 and 4.

Several teacher-level characteristics were associated with changes in SGP effectiveness. Traditionally certified teachers had greater growth which resulted in higher SGP scores at the end of the year, $M = 52.82$, $SD = 13.06$, when compared to their alternatively certified counterparts, $M = 49.93$, $SD = 12.76$, $p < .001$, and this coefficient remains significant even with the inclusion of school effects. Though it was not one of the central research questions, this analysis provides evidence in a contested debate among scholars about the relative merit of traditionally certified vs. alternatively certified routes into teaching (Darling-Hammond, Berry, & Thoreson, 2001; Goldhaber & Brewer, 2000). Notably, Aldine ISD has a fairly strategic recruitment process, where they rely on attracting candidates from highly ranked teacher education schools to do their student teaching in Aldine and then encourage those student teachers to move into full-time positions. This may help explain the significant results. Additionally teachers' years of experience were negatively correlated with SGP effectiveness ($B = -.06$); however, this coefficient was no longer statistically significant when including school fixed effects, as shown in Model 4. Unlike the case with motivation, the inclusion of school effects

impacts the overall estimates, indicating the importance of school-level characteristics (which will be explored in more detail in Chapter 6).

Table 4-6

Pilot's Impact on Teacher Effectiveness (as measured by SGPs)

Measure	Without School Effects		With School Effects	
	Model 1 *Impact of the Pilot*	Model 2 *With Individual Controls*	Model 3 *Impact of the Pilot*	Model 4 *With Individual Controls*
Impact of INVEST (Pilot*Year)	-0.06 (0.10)	0.01 (0.11)	-0.03(0.11)	0.02(0.10)
Pilot	0.08 (0.10)	0.05 (0.11)	0.53(0.10)***	0.34(0.14)*
Year	0.06 (0.07)	0.05 (0.07)	0.07(0.08)	0.08(0.07)
Gender (female)		-0.04 (0.07)		0.02(0.07)
Ethnicity (white)		0.04 (.08)		0.06(0.06)
Certification (traditional)		0.24 (0.06)***		0.26(0.02)***
Years of Experience		-0.06 (0.04)*		-0.04 (0.04)

Note. N = 906. All continuous variables have been standardized.
**p < .05. **p < .01. ***p < .001*

As discussed in Chapter 1, a growing body of research has examined the validity and reliability of new teacher effectiveness measures. Historically, researchers have found measuring teachers' performance to be incredibly challenging given the reality that quality teaching is influenced by contextual factors and further, that there is considerable disagreement about the validity of various outcomes (e.g., test scores). To address this concern, INVEST mirrors most new performance management systems and uses multiple indicators, which raises questions about the correlation between evaluation measures. As demonstrated in Table 4-7, there is a significant but relatively low correlation between SGPs and Danielson, $r = .27$, $p < .001$, which is consistent with other recent research

(Kane, McCaffrey, Miller, & Staiger, 2013). Given the fact that the Danielson Framework was used as part of the system during the pilot year, it is likely the better indicator of teacher performance at this stage. However, since it was not available prior to the start of the pilot, it unfortunately could not be used in impact analysis for this year.

Regardless of the effectiveness measure used, the impact on teachers' improvement may take longer than a year to translate into changes in practice (particularly given implementation challenges associated with the pilot). In support of this postulation, Table 4-7 shows that teachers' reported change in practice (on the survey) was not significantly correlated with their performance on either SGPs (r = -.12) or Danielson scores (r = .04) in the pilot year. Initially, I planned to assess whether teachers' motivation mediated teachers' level of effectiveness, but since the pilot did not impact teacher effectiveness, I was not able to test for mediation. However, as demonstrated in Table 4-7, teachers' level of motivation (both personal and system) was associated with their Danielson rating at the end of the pilot year. Teachers' personal value for teaching, as well as their value for the system, were both positively correlated with the Danielson rating, r = .14, p < .001, though this is relatively small in magnitude and does not indicate the directionality in the relationship. In other words, receiving a higher rating could have increased teachers' value or teachers' value could have led them to perform better on the system. Though both personal and system expectancy are correlated with the Danielson rating, these relationships are quite small (r = .08 and r = .11, p < .001 respectively); however the relationship between system expectancy and reported changes in practice is a considerably greater, r = .25, p < .001. These

correlations suggest that there is a positive (and statistically significant) relationship between teachers' level of motivation, their reported changes in practice, and their effectiveness (as measured on the Danielson Framework). Interestingly, none of these motivational measures were correlated with teachers' performance on the Student Growth Percentile measure. This finding could be due to limitations in the Student Growth Percentile measure since it only captures one year of data. Alternatively, since teachers were not made aware of their results on SGPs, we might also expect a lagged effect and should examine whether these results are significant in a subsequent year.

Table 4-7

Correlation between Effectiveness Measures

Measure	1.	2.	3.	4.	5.	6.	7.
1. SGPs (end of year)	--	0.27***	0.08	-0.12	0.04	-0.01	0.06
2. Danielson (end of year)		--	0.04	0.08**	0.11***	0.14***	0.14***
3. Changes in Practice			--	0 .15***	0.25***	0.20***	0.18***
4. Personal Expectancy				--	0.24***	0.46***	0.26***
5. System Expectancy					--	0.23***	0.26***
6. Personal Value						--	0.27***
7. System Value							--

**p<.05. **p<.01. ***p<.001*

Qualitative Results

There are several trends in the qualitative data that help elucidate the quantitative analysis described above. As discussed in Chapter 1, even if teachers are initially motivated to improve performance, when individuals move from the deliberation to the implementation phase, they need to commit to a specific goal and develop a plan for

translating that goal into action. To do so, individuals must engage in deliberate practice by setting specific interim goals, receiving immediate feedback on performance, and consistently working towards mastery of internalized goals. As discussed in Chapter 3, since INVEST utilized a new online system, most teachers initially anticipated receiving detailed feedback and engaging in more frequent conversations about improving their practice. Despite the presence of these systems and structures, principals were overwhelmed by the timeline in the early months of the year and while some schools provided extensive support for teachers, many schools struggled to implement the new system with fidelity. As a result, teachers reported that they did not receive adequate feedback to improve performance over the course of the year. This was particularly frustrating for teachers who were no longer at the top of the evaluation system and came to believe that no matter how hard they worked, Level 4 performance would remain unreachable.

Even when teachers reported receiving the required number of observations, they questioned whether the evaluation process as a whole was set up to provide feedback that could result in meaningful changes in practice. When fully implemented, veteran teachers received three walkthroughs and one formal observation and novices received six walkthroughs and two formals over the course of the year (with at least a month typically between observations). One teacher noted how more immediate feedback would have helped her determine if her instruction was effectively meeting students' needs: *"So if you are going to come and observe a lesson one day, then to see whether the changes worked or not, you need to come back the next day…and help me determine what to*

improve" (School 2, Teacher 6). Given administrator time constraints, this type of immediate and targeted feedback was a fairly rare occurrence. Additionally, though teachers reported they had a clear picture of where they were performing in relation to the standards (given the specificity of the Danielson Framework), the feedback they received on how to improve performance was not necessarily actionable. Teachers needed a clearer picture of what the next level of performance looked like in practice and more concrete and specific steps to take to become Level 4 teachers. Though the Teachscape modules assisted by providing videos of various components, teachers expressed concern that these videos *"weren't grade-level appropriate"* or *"of our kids,"* and as a result, did not reflect their experience as a teacher in the district.

Theory suggests that goals are most likely to be achieved when individuals are engaged in self-reflection and planning. Practically speaking, given the constraints on administrators' time, the evaluation process appeared to be more effective when teachers were able to drive their improvement efforts. For some teachers, INVEST pushed them to self-reflect and changed the way they approached their teaching: *"I've had to engage in self-reflection which has made me a better teacher. It's made me a more thoughtful teacher as far as the lessons that I'm delivering, how I'm delivering them and how I'm planning the interaction around what we're doing"* (Teacher 3, School 1). These teachers owned their own improvement process; as one such teacher remarked, "*I'm one of those people who puts it on myself. It's something I need to figure out"* (Teacher 3, School 4). To the contrary, when feedback was viewed as too prescriptive, it had the adverse effect and reduced the likelihood that it would result in changes in practice. As one teacher

shared, *"I understand they have to have rubrics but sometimes it's like all the teachers we don't all fit in one box. Sometimes we have to deviate from what the rubric says. As teachers we need to do what's best for our kids"* (School 2, Teacher 3). Though there was considerable variation in the feedback process (which will be discussed in Chapters 6 and 7), it was more common for teachers to report that it was principal-driven than focused on building teacher ownership. As a result, it is perhaps not surprising that INVEST did not have an overall positive impact on teachers' effectiveness (as measured by SGPs) during the pilot year.

Section Three: Retention

To sustain improvements over time and commitment to the profession itself, teachers must stay engaged and avoid experiencing burnout. As theory demonstrates, burnout is characterized by a state of exhaustion where an individual becomes cynical about their value and impact in their work, resulting in higher rates of occupational turnover. It is characterized by three components – (1) emotional exhaustion, which measures feelings of being emotionally overextended and exhausted by one's work, (2) depersonalization, which measures an unfeeling and impersonal response towards the beneficiaries of one's instruction (similar to a lack of value), and (3) lack of personal accomplishment, which measures feelings of competence and successful achievement (similar to a lack of expectancy). To measure the systems' impact on burnout, teachers responded to the burnout scale on the end of year survey, and then teacher turnover was tracked at the individual level the summer following the implementation of the pilot.

It is important to note that not all teacher turnover was viewed as negative by the district. Indeed, one of the goals of INVEST was to remove ineffective educators from the classroom. Thus, in addition to describing retention in the aggregate, this chapter will also examine the level of teacher burnout and turnover by effectiveness level.

Quantitative Results

Descriptive Evidence

Table 4-8 captures the unadjusted trends in teacher turnover and burnout for all Aldine ISD schools, the pilot schools, and the non-pilot schools. The year prior to the implementation of INVEST (in spring 2012), teachers in non-pilot schools had higher percentages of teachers leaving the district (M = 10.26, SD = 4.63) when compared to teachers in pilot schools (M = 9.75, SD = 4.30). After the implementation of the new system, turnover rates increased in both pilot and non-pilot schools, but grew more significantly in pilot schools (+3.42), compared to non-pilot schools (+2.00). In terms of burnout, which is one possible indicator of future turnover, teachers' burnout in pilot schools was slightly higher in pilot schools (M = 2.98, SD = .93) compared to non-pilot schools (M = 2.83, SD = .33).

Table 4-8

Descriptive Statistics: School-Level Turnover and Teacher-Level Self-Reported Experiences

Measure	**Pilot**	**Non-Pilot**
School-Level Turnover (previous year)	9.75 (4.30)	10.26 (4.63)
School-Level Turnover (end of year)	13.17 (5.20)	12.26 (5.05)
Teacher Burnout *(captured from end of year survey data, scale 1-5)*	2.98 (0.927)	2.83 (0.33)
Teacher Turnover Intentions *(captured from end of year survey data, Scale 1-5)*	2.33 (0.27)	2.25 (0.33)

Note. N = 74 schools since turnover can only be collected historically in the aggregate at school level

Impact Estimates

Using the difference-in-differences approach, I estimated the effect of the evaluation pilot on teachers' turnover, which only included teachers who left the district entirely (leavers). Unlike the motivation and effectiveness measures, turnover data pre- and post-pilot implementation was only available at the school level, so I ran three instead of four models. The first model did not include individual or school level covariates, meaning it did not account for individual or school characteristics that may have contributed to observed outcomes in teacher turnover. The second model included individual covariates and the third model included individual and school covariates to create a more robust estimate of the pilot's impact on teacher turnover.

As demonstrated in Table 4-9, the pilot had a positive impact on teacher turnover (or in other words, a negative impact on teacher retention). The relative change in pilot schools' turnover rates at the end of the year (the "difference in differences") was 0.29

standard deviation units (or 1.42 in unstandardized units) though this difference was not statistically significant in any of three models. It is important to note that since this analysis was run at the school-level, the small sample size (N = 34 for pilot schools and N = 40 for non-pilot schools) may be contributing to the lack of statistical significance. As demonstrated in Models 2 and 3, the magnitude of this estimate is not affected by the inclusion of individual or school covariates.

Teacher turnover differed across teachers with various characteristics. More experienced and white teachers had slightly lower rates of turnover then their less experienced and non-white colleagues; however, the difference in ethnicity was not statistically significant when controlling for school effects. The influence of teacher experience persisted even when including school controls, and an increase in one standard deviation in teachers' years of experience was associated with a decrease in .11 standard deviations in the rate of teacher turnover, $p < .001$. Novice teachers had higher rates of teacher turnover than their experienced counterparts. Indeed, 13.29% of teachers in their first three years in this dataset left teaching in the district at the end of the 2012-2013 year, compared to 9.80% of teachers with more than three years of experience, and this turnover rate was close to 15% for teachers in their first year teaching in pilot schools.

Though teacher effectiveness data could not be included in this school-level model, descriptive data suggested that the preliminary results trend in the direction of the district's theory of action. The pilot schools had higher turnover rates of Ineffective Teachers (12.50%) and Needs Improvement Teachers (12.63%) when compared to

Effective (10.13%) and Highly Effective Teachers (9.00%) on the Danielson Framework for Teaching. In other words, though turnover rates increased slightly in pilot schools, this increase appeared to be due to the exiting of a greater percentage of underperforming teachers. However, it is important to note that without prior years of data, we cannot know whether this was a continuation of previous years' trends or a result of the INVEST pilot. Unlike the Danielson Framework for Teaching, there were no notable differences in the turnover rates of teachers based on their performance on Student Growth Percentiles. As noted above, this is likely due to the fact that the majority of teachers had yet to be made aware of their performance on the student growth metric.

Notably, none of the school level controls were associated with teachers' level of turnover. This finding was somewhat surprising given prior research on the influence of school characteristics on rates of teacher turnover (Ingersoll, 2001). However, the demographics of Aldine ISD help shed some light on this initially striking finding. As noted in Chapter 2, more than 84.9% of all Aldine students are classified as economically disadvantaged and receive Title I support and the racial composition is 95.9% non-white, so there is very little variation in these measures across campuses. As far as enrollment, the district is also unique in that it has quite large enrollment across different levels of schooling. So while there are considerably more students at the high school level (M = 1810) when compared to the Pre-K level (M = 657), there is very little variation in enrollment size across the other levels of schooling (average sizes of other levels of schooling range from 819 to 947). This may help to explain why enrollment does not predict teacher turnover at the school level.

Table 4-9

Pilot's Impact on School-Level Turnover

Measure	Model 1 *Impact of the Pilot*	Model 2 *With Individual Controls*	Model 2 *With Additional School Controls*
Impact of INVEST (Pilot*Year)	0.29 (0.32)	0.26 (0.31)	0.29 (0.32)
Pilot	-0.10 (0.21)	0.04 (0.21)	-0.00 (0.22)
Year	0.40 (0.22)	0.38 (0.21)	0.43 (0.24)
Gender *(female)*		-0.02 (0.01)*	0.00 (0.01)
Ethnicity *(white)*		0.01 (0.01)	-0.01 (0.01)
Certification *(traditional)*		0.01 (0.01)	-0.01 (0.01)
Years of Experience		-0.11 (0.04)**	-0.13 (0.04)**
Percent African American			0.02 (0.04)
Percent Hispanic			0.04 (0.04)
Economically disadvantaged			-0.02 (0.02)
Limited English Proficient			-0.01 (0.01)
Enrollment			0.01 0(.00)
Performance *(Student Growth Percentiles)*			-0.01 (0.01)

Note. $N = 74$ (School-Level Analysis). All continuous variables have been standardized. *$p < .05$. **$p < .01$. ***$p < .001$

Teacher turnover is influenced by a number of different factors, which will be discussed in more detail in Part 2. However, one cause of turnover which has the potential to result from new and demanding evaluation systems is employee burnout. As demonstrated in Table 4-10, teachers' burnout was significantly correlated with their turnover intentions (i.e., whether they reported planning to leave teaching), $r = .45$, $p < .001$ and to a lesser extent with actual turnover (i.e., whether they left the district in the pilot year), $r = .08$, $p < .001$. As reported in Table 4-9 above, teachers' higher rates of

burnout in pilot schools suggest that burnout may be a potential concern to monitor for future years of the system.

Teachers' level of burnout and turnover intentions were negatively correlated with teachers' value and expectancy (both personal and system). Overall, teachers' personal value for teaching, as well as their personal expectancy, were negatively associated with burnout, which is r = -.41 (for value) and r = -.34 (for expectancy), both $p < .001$, and desire to leave teaching, r = -.38 (for value) and r = -.19 (for expectancy), both $p < .001$.

Table 4-10

Correlation between Teacher-Level Turnover, Burnout, Turnover Intentions, and Motivation

Measure	1.	2.	3.	4.	5.	6.	7.
1. Turnover *(teacher-level)*	--	0.33***	0.08***	-0.02	0.01	-0.04*	-0.03
2. Turnover Intentions *(end of year survey)*		--	0.44***	-0.19***	-0.20***	-0.38***	-0.23***
3. Burnout *(end of year survey)*			--	-0.34***	-0.25***	-0.41***	-0.20***
4. Personal Expectancy *(end of year survey)*				--	0.24***	0.46***	0.26***
5. System Expectancy *(end of year survey)*					--	0.23***	0.26***
6. Personal Value *(end of year survey)*						--	0.27***
7. System Value *(end of year survey)*							--

**p<.05. **p<.01. ***p<.001*

Teachers' overall motivation for teaching was more associated with burnout and turnover intentions than their motivation on the system itself; however, both system value and system expectancy were also negatively correlated with burnout and turnover intentions. In sum, these correlations suggest that more motivated teachers were less likely to

burnout and leave the district. The qualitative data help provide a more in depth understanding of these quantitative trends.

Qualitative Results

Most teachers (as all new employees) began teaching engaged in their practice and then for some, certain experiences caused this initial enthusiasm to turn into cynicism and perceptions of ineffectiveness. To avoid this type of burnout, teachers must continue to place value on their work and believe in their expectancies as educators. By focusing attention on teachers' impact on student growth, INVEST had the potential to create a more personalized connection between teachers and their students and subsequently, to build value in their work. Indeed, at the beginning of the year, many teachers reported that they appreciated the fact that the rubric was "student-centered" and believed it would make them more focused on building relationships with their students.

However, by the end of the year, INVEST had resulted in burnout for a subgroup of teachers. In particular, INVEST lowered expectancy for teachers who did not reach Level 4 performance, which challenged their perceptions of competence and led to heightened frustration. This type of experience with burnout eroded teachers' professional identify, making them feel unappreciated for their hard work. Several teachers speculated that these feelings of frustration and exhaustion could lead to higher rates of teacher turnover:

> The morale is going to go down and the people will leave the district and go somewhere else that doesn't have INVEST. INVEST is a major problem because a lot of people here are already looking at other districts to get out of Aldine. They don't want to stay in Aldine. And it's INVEST. Because at first they were like, oh, OK, let's see what's going to happen and then when we got toward the end, like now, and everyone's getting their summatives, and even when they were

> getting their walkthroughs throughout the year, they weren't in agreement with it, so I think people are not just going to go to different schools, they are going to leave the district. Like I said, the morale is horrible (School 3, Teacher 1).

As discussed above, this lowered expectancy was exacerbated by the terminology associated with the new system. Many of the veteran educators who had become conditioned to performing well on the PDAS evaluation system found it insulting to be referred to as merely "effective" or "proficient." This new terminology did not recognize teachers' accomplishments, which may have contributed to increased turnover (or may do so in subsequent years), particularly among veteran teachers.

In addition to feeling frustrated by not meeting expectations, many teachers were overwhelmed by the increased expectations associated with INVEST (e.g., watching modules on the Danielson Framework, assembling an artifact binder with results from the year). INVEST increased expectations for teachers without providing additional time outside of their other responsibilities, and the demands placed on teachers often outstretched existing capacity. In particular, INVEST created additional paperwork, which required additional time from teachers but did not always align with their valued outcomes. As one teacher shared, *"We have to do all the documentation turn in all the proof... It's just more work on the teacher. I am so buried in paperwork which means by teaching falls by the wayside"* (School 6, Teacher 2). This frustration, coupled with the imbalance of demands and resources, left a subgroup of teachers feeling burnt out from teaching in the district.

However, it is again important to note that these concerns over burnout disproportionality affected lower-performing teachers. Indeed, teachers performing at the

Highly Effective level on Danielson had significantly lower levels of burnout comparatively. This analysis provides preliminary data that suggests the policy may have promoted what some policymakers have referred to as "strategic retention." In other words, the system appears to do a better job of retaining the high performers and encouraging the low performers to exit the district.

Summary

In sum, this chapter demonstrated that INVEST did not have an overall positive impact on teachers' outcomes of interest. For motivation, INVEST had no impact on teachers' personal value (or belief in the importance of their work) and a negative impact on teachers' personal expectancy (or belief in their own ability as educators). Notably, teachers' personal motivation for their work was only moderately correlated with their motivation for the INVEST system itself. At the beginning of the year, teachers generally reported that the Danielson Framework offered clear and specific expectations, which helped build initial personal – and system – expectancy. However, over the course of the system's implementation, frustration heightened among teachers who were unable to meet the Level 4 expectation of creating student-led classrooms, subsequently lowering both personal and system expectancy. Teachers' motivation to stay committed in the face of these challenging new standards was influenced by their value for the system. If teachers felt the system aligned to their own vision of effective teaching and would support their growth as professionals, they were more likely to value performing well on INVEST. However, if they viewed the system as creating an unnecessary additional workload, these costs contributed to lower overall value.

Perhaps not surprisingly, given the limited effect on motivation, INVEST did not have a significant impact on teacher effectiveness (as measured by the Student Growth Percentile metric). Since the Danielson Framework for Teaching (observation) measure could not be analyzed longitudinally due to a lack of baseline data, this analysis relied solely on the Student Growth Percentile (SGP) measure to capture effectiveness. Though the two measures were significantly correlated, the size of the relationship was relatively small, $r = .27$, $p < .001$. Unfortunately, the SGP measure was limited because it only applied to teachers in tested subjects and the majority of principals did not share this data with their teachers during the pilot year. As discussed in Chapter 1, to increase effectiveness, teachers need to engage in a process of deliberate practice, where they monitor their progress towards goals and receive immediate, specific, and consistent feedback throughout the course of the year. Since the SGP data was not made available to teachers, this was not the kind of metric they could use to actively improve performance. In contrast, teachers could use the Danielson Framework during the year to set interim goals and receive feedback on specific aspects of practice. However, due to challenges with implementation, there was variation in the extent to which teachers reported receiving targeted support from their principals and making changes in their practice.

The impact on teacher retention was more nuanced. Overall, pilot schools had slightly higher turnover rates than non-pilot schools, though these differences were not statistically significant. Qualitative data suggested that this increase in turnover in pilot schools seemed to be driven in part by the same frustration (and burnout) resulting from the perceived unattainability of Level 4 performance on the new evaluation system

coupled with an intensified workload. However, these rates were not consistent across all subgroups of teachers. In particular, Highly Effective teachers appeared less likely to feel burnt out by their work and more likely to stay at their schools when compared to their Ineffective and Needs Improvement counterparts, which suggests that the policy may indeed be having its intended effect on teacher retention by retaining the higher performers and encouraging the lower performers to leave the district.

While informative, these overall trends only begin to tell the story of INVEST's implementation and impact. Given the limited or null effects, one might conclude that INVEST did not have much influence on teachers in either direction. In fact, many of the previous studies on performance management systems have stopped at this stage of the analysis and reached similar conclusions. Regrettably, these studies leave much to be told about the realities of implementation. In Aldine ISD, overall trends masked considerable variation in teachers' responses to INVEST. As was the case with teacher retention, some of this variation across outcomes was associated with teachers' individual characteristics. Additionally, as discussed in Chapter 3, this variation was also driven by teachers' perceptions of specific design features of INVEST. Both individual and system level factors were influenced by the climate and fidelity of implementation at the school level. Indeed, despite these overall trends, this analysis revealed that all three outcomes of interest – motivation, effectiveness, and retention – were influenced by teachers' individual characteristics, school organizational factors, and perceptions of system features. Each of these sets of factors will be examined in turn in part two of this dissertation.

PART TWO FINDINGS: VARIATION

The trends presented in Chapter 4 suggested that, overall, INVEST did not have a positive impact on teachers' outcomes. In fact, the system had a slightly negative impact on teachers' expectancy or belief in their ability to improve their practice and no impact on the value they placed on their work or their effectiveness in improving student growth. Though there was no significant impact on teachers' retention overall, some preliminary descriptive evidence seems to suggest that the system may have been successful in exiting lower-performing teachers from the district. However, these overall results are incomplete, as they do not capture the considerable variation in each of the three outcomes of interest. As theory would predict, teachers' attitudes and behaviors were influenced by their perceptions of how well the system was designed and implemented, as well as by their individual differences and the contexts in which they worked.

This second part of the dissertation elucidates the results presented in Chapter 4 by exploring how outcomes were influenced by variation in individual, school, and system characteristics. I used multiple regression analyses to examine which factors (at the individual, school, and system level) predicted outcomes of interest – i.e., teacher motivation, effectiveness, and retention. For each outcome, I ran several regression analyses, where I entered a new set of predictors (i.e., individual characteristics, school characteristics, system characteristics) sequentially to determine the relative impact of various types of variables. This quantitative data was supplemented with qualitative analysis (from interview data) of how variation across individual characteristics, school

organizational factors, and system design features influenced teachers' motivational responses and performance on the new system. Given the interest in the impact and implementation of INVEST, analysis is limited to data collected on teachers in pilot schools (N = 1097 for quantitative data and N = 36 for qualitative data) and will primarily explore how various predictors influenced teachers' motivation and performance on the system itself. Correlational results are presented for the Student Growth Percentile outcome measure as well, but the focus will be on the Danielson Framework measure since this was shared with teachers during the pilot year of implementation. The analysis is divided into three chapters to explore each of the sources of variation.

Chapter 5: Individual-level variation. As discussed in Chapter 3, though the most common teacher survey response regarding the features of INVEST was "neutral," teachers' attitudes differed significantly across subgroups, such as by years of experience and effectiveness level. Indeed, both the qualitative and quantitative data revealed considerable variation at the individual level. For veteran teachers, INVEST was a substantial departure from their typical experience with evaluation. As would be expected with any meaningful change, individuals responded quite differently. In this chapter, I use the qualitative data to categorize teachers' responses into five distinct profiles and then explore how teachers' personality characteristics (captured on the teacher survey) predicted outcomes.

***Chapter 6: School-level variation*.** Chapter 3 demonstrated that teachers had high expectations for the system's promise; however, INVEST did not live up to its potential for many teachers due to challenges with the fidelity of implementation. High school

teachers and teachers in higher performing schools tended to have less positive attitudes towards the new system. My sampling strategy for school case studies captured variation across levels (e.g., elementary, high) and school performance levels (e.g., both higher-performing and lower-performing schools). In addition to varying along level and performance, these school case studies diverged along school climate indicators. In this chapter, I describe implementation at the six case study schools and then investigate how teachers' perceptions of school climate (captured on the teacher survey) predicted outcomes.

Chapter 7: System-level variation. In Chapter 3, I explored evaluation attitudes (more generally) and then attitudes toward the new system (INVEST specifically). Analysis demonstrated that at the end of the year, teachers in pilot schools had lower perceptions of evaluation when compared to teachers in non-pilot schools across all survey metrics. Certain design features of the new INVEST system (i.e., level of understanding of the purpose of the new system, perceived accuracy of the new evaluation measures, and the quality of feedback and support) appeared to shape teachers' responses to the new policy. In this chapter, I explore how these design features influenced teachers' outcomes on the new system.

CHAPTER 5: INDIVIDUAL-LEVEL VARIATION

To influence teacher motivation and behavioral change, theory indicates that evaluation systems must align with the preferences of the individuals they are designed to impact. Given this, we would expect motivational responses to vary across subgroups of teachers – in particular, by years of experience, effectiveness, and personality. Preliminary evidence from Chapter 3 suggested that newer teachers, as well as highly effective teachers, were more likely to have positive perceptions of INVEST, and thus, appeared to be more open to making changes in their practice. Additionally, research has demonstrated that individual differences in teacher personality shape teachers' experiences (Somech, 2010; Teven, 2007) and that certain individuals will be more predisposed to sustain improvements in practice over time (Achtziger & Gollwitzer, 2010). In this chapter, I explore this individual variation by first categorizing teachers' responses into five distinct profiles (Section One) and then examining which personality characteristics were most predictive of outcomes under INVEST (Section Two).

- ***Section One: Teacher Personality Profiles.*** In my interviews, I was struck by teachers' varied attitudes towards INVEST and how they processed their experiences with the new system. Though no two teachers responded identically, several trends emerged from the data that helped me better understand and ultimately categorize teachers' responses. As discussed in Chapter 2, I used the codes I developed in Atlas.ti to classify teachers into one of five distinct personality profiles – invested teachers, sponge teachers, burnt-out teachers, insulted teachers, and skeptical teachers. The majority of teachers I interviewed

could be easily classified into one of these five profiles. However, several teachers possessed attributes of multiple profiles. For these teachers, I worked with two research assistants on the project to discuss the best placement for each teacher based on the majority of their responses.

- ***Section Two: Key Individual Characteristics.*** After exploring these personality profiles, I then set out to systematically investigate how personality influenced outcomes within the larger population of teachers in the district. Although many personality inventories exist, the five-factor theory has emerged as the foundational approach to describing personality traits (Goldberg, 1990**;** John & Srivastava, 1999; McCrae & Costa, 1987), so I employed a shortened and previously validated version of this inventory in my teacher survey. I also used two measures of grit, defined as passion and perseverance for long-term outcomes – a teaching-specific measure I developed and an overall domain-general measure that had been previously validated and used with teachers in a prior analysis (Duckworth et al., 2009). In this section, I examine how the teacher personality scales influenced teachers' motivation, effectiveness, and retention.

Section One: Teacher Personality Profiles

This section provides an overview of the five personality profiles I developed from the qualitative interview data – invested teachers, sponge teachers, burnt-out teachers, insulted teachers, and skeptical teachers. Table 5-1 summarizes the key personal attributes and overall reactions for each of these five types. In the Appendix, I provide additional information on the survey responses for those teachers I interviewed that also

completed the survey. As these data make clear, teachers' responses to INVEST were notably quite distinct across each of these profiles.

Table 5-1

Teacher Profiles Types (from Interview Data)

Teacher Profiles	Personality Description	Overall Reactions to INVEST
Invested Teachers *N = 5*	Invested teachers were veterans with over 10 years of experience (mostly within the district). They were ***open*** to new experiences, displayed ***grit*** by engaging in strategic goal-setting, and remained ***conscientious*** in working towards desired outcomes.	Invested teachers viewed challenges associated with the new system as within their control and set rigorous performance goals on INVEST. They were open to feedback and took ownership over implementing changes in their practice. Their motivations were primarily intrinsically driven and they valued the impact they had on students.
Sponge Teachers *N = 6*	Sponge teachers were novices with a similar profile to invested teachers. They were very ***open*** and ***conscientious*** which was driven by their ***agreeable*** nature and strong desire to prove themselves as new teachers. Though they were hard working, they struggled to set long-term goals, given their inexperience.	Sponge teachers were very open to the new system. They wanted to perform well on INVEST, which was partially driven by an extrinsic desire to please the administration. As the term "sponge" would suggest, they readily absorbed feedback and were quick to implement changes in practice as a result of the new system.
Burnt-out Teachers *N = 6*	Though they were primarily new teachers, burnt-out teachers were less ***open*** and ***agreeable*** than their sponge counterparts. They were fairly ***conscientious*** in implementing aspects of the new system, but they lacked the ***grit*** necessary to stay focused on long-term	Burnt-out teachers were overwhelmed by the new system and struggled to keep up with the requirements of teaching in the district. INVEST was not necessarily the source of their frustration but it contributed to their overall level of exhaustion. They did not see

	goals and respond effectively to challenges.	the value of the new system and instead viewed it as merely a source of additional work.
Insulted Teachers *N = 6*	Insulted teachers were ***conscientious*** and reported being very ***gritty*** and focused on achieving highly effective status on the new evaluation system. As a result, when they did not reach high standards, they were incredibly frustrated which made them less open to change and more negative in general.	Insulted teachers were incredibly frustrated by the new system. Their reactions were primarily driven by the perceived unfairness of the new evaluation measures (in particular, the unattainability of Level 4 performance). Instead of viewing INVEST as an opportunity to take their practice to the next level (like invested teachers), they questioned the new system's intentions.
Skeptical Teachers *N =13*	Skeptical teachers were primarily veteran teachers who were generally ***agreeable*** in nature. However, they were very ***practical and analytical*** individuals, which made them initially more ***neurotic*** in response to change. Some skeptical veterans were entirely ***apathetic*** while others were just less likely to embrace reform until they had seen how it played out in practice.	Unlike other groups, skeptical teachers did not have strong reactions to INVEST and remained fairly neutral toward the system over the course of the year. They were quite confident in their abilities but were not as invested in the new system and were less optimistic about whether INVEST could have an impact on performance. Given their (often negative) experience with previous systems, they raised very specific and technical questions.

Invested Teacher Profile

Overall, the group of invested teachers was extremely positive about their teaching experience and the new evaluation system. Not only did they start the year off with a receptive attitude, but invested teachers also remained optimistic throughout the

entire year and viewed challenges as a necessary obstacle to overcome rather than as external to their control. For the most part, invested teachers were very confident in their teaching ability and viewed their students' performance as a reflection of their own actions. As a result, they remained open to feedback and committed to doing whatever they perceived it would take to reach their goals. There were five invested teachers, and three of them were in School 1 (the other two in Schools 3 and 6). Given their longer-term commitment and perspective, all five of the invested teachers were veteran teachers, and they all had over 10 years of experience in teaching in the district.

Invested teachers had a distinct personality profile – they were open to new experiences (like INVEST), displayed grit by engaging in strategic goal-setting, and remained conscientious in working towards desired outcomes. Though they had questions about INVEST, invested teachers were open to and excited by the possible impact the new system would have on their practice. One invested teacher shared, *"I was actually looking forward to it because it was different from our old system and would provide me feedback on how to grow"* (School 1, Teacher 3). For them, INVEST was a tool that could be used to help accelerate their growth as educators.

Given their investment in the system, these teachers all desired to reach the top level of performance and consequently set specific goals for their improvement. For the teachers who did not reach their goals this year, they were confident in their abilities and expressed an ongoing desire to continue to improve their practice. One invested teacher shared, *"I'm sure I'll get to level 4 in a couple more domains…I'll always have something to work on, to make better"* (School 6, Teacher 1). These teachers stayed

motivated throughout the year and went above and beyond to reach their goals. As one described, *"I need to come to school early because I need to be 100% focused on students"* (School 1, Teacher 6). Indeed, they were very conscientious and recognized the importance of meeting deadlines and staying on top of all relevant paperwork.

Sponge Teacher Profile

Overall, the group of sponge teachers was extremely open to the new system because (as entirely first year teachers), INVEST was the only evaluation system they had experienced. As several sponges shared, *"we didn't know any differently. INVEST was all we've had."* Similar to invested teachers, sponge teachers desired to improve their instruction; however, their desire to perform well was not only intrinsically driven, but also appeared to be a result of their desire to be recognized by the administration. As first year teachers, they felt they had something to prove and this attainment value influenced their motivation on the new system. There were six sponge teachers in five of the six schools, all in their first year in the classroom.

Sponge teachers had a similar personality profile to invested teachers – they were open to new experiences (like INVEST) and remained conscientious in working to improve their practice over the course of the year. However, since sponges were first year teachers, they had a more limited ability to set and work strategically towards long-term goals. Nonetheless, they did express a strong desire to improve their practice towards specific outcomes and readily absorbed feedback. One sponge teacher shared, "*I was open to it because as a new teacher I do want feedback on how I can become a better teacher especially to make it more student centered because that is my goal"* (School 3,

Teacher 4). Several of the sponge teachers were so committed to improvement that teaching had consumed their lives. One shared, *"I'm going to do everything all the time. I kind of have this motto with teaching, there's always something I could be or should be doing."* (School 4, Teacher 3). Sponge teachers had particularly high levels of agreeableness, which meant their desire to perform well on the system also seemed to be driven by a desire to be perceived by others as competent. They spoke frequently of their interactions with colleagues and how important it was for them to have a strong relationship with their mentors and administration. Though the workload was challenging, they remained conscientious and recognized that extra work was required to meet valued expectations.

Burnt-out Teacher Profile

Overall, the group of burnt-out teachers was overwhelmed by INVEST and generally had more negative reactions to the system as a whole. They were all first year teachers in the district (though some had teaching experience in other districts). Their negativity stemmed from a general exhaustion associated with the challenges of teaching, more so than from specific aspects of the new system. Indeed, INVEST was not the source of their frustration, but only exacerbated challenges; as one burnt-out teacher shared, *"INVEST was just more work. The work is just never-ending. And it's one more thing"* (School 6, Teacher 3). There were six burnt-out teachers, one in each of five schools (with the exception of School 4 where the new teachers were primarily sponge teachers).

Burnt-out teachers had very different personality profiles than invested and sponge teachers – they were more neutral to new experiences (like INVEST) and though they were fairly conscientious, they were often overwhelmed by the challenges associated with teaching in the district. One burnt-out teacher shared, *"like I said, this whole thing has really just been a thorn in my side"* (School 1, Teacher 5). In their mind, the expectations associated with INVEST were not only unnecessary, they were also unrealistic. Unlike sponge teachers, burnt-out teachers did not have particularly high levels of agreeableness, which meant they were not as concerned about raising concerns with the new system. One burnt-out teacher shared that she had no problem bringing her frustration to the administration, *"I have all of these things and it's like none of that is being taken into consideration. I'm just supposed to keep going like the Energizer bunny. Just keep going and going and going. I'm not willing to do it"* (School 1, Teacher 5). Though they generally valued feedback on their performance, burnt-out teachers did not always find the system's goals to be aligned with their own. Perhaps somewhat surprisingly, the burnt-out teachers all scored in the effective range and outperformed some of the sponge teachers at the same school. Indeed, their different perceptions were not driven by their performance, but rather by their perceptions of the feasibility of the expectations. As this burnt-out teacher makes clear, teachers were being asked to do the impossible: *"Even next year as a second year teacher, it's not going to happen. There's no way I can get a 4"* (School 4, Teacher 5).

Insulted Teacher Profile

Overall, the group of insulted teachers was incredibly frustrated by the new INVEST system, and their reactions were driven by their perceptions of unfairness and overall frustration with the expectations. The majority of insulted teachers had achieved the top level of performance on the old evaluation system, PDAS, but did not reach that level on INVEST. Instead of looking at the new system as an opportunity to further develop their practice, like invested teachers had, insulted teachers viewed the expectations as unfair and felt the system devalued their work. In total, there were six insulted teachers in five of the six schools, and all but one had multiple years of teaching experience.

Insulted teachers had similar personality profiles to burnt-out teachers, though their lack of openness to the system was driven by frustration rather than exhaustion. Two of the insulted teachers were not open to the system from the beginning of the year and showed a lack of emotional stability (or neuroticism) by assuming the system was designed to demean teachers' work. One of these teachers shared, *"it's given the administration like basically a way to blackmail teachers into doing all this ridiculous, ludicrous garbage with the threat hanging over their heads. It's so completely not acceptable"* (School 6, Teacher 2). These insulted teachers appeared to be quite conscientious in meeting expectations but resented the additional workload. The same insulted teacher went on to share, *"INVEST is being used as a club against teachers, as a bullying tactic, as a weapon to get us to do more than we should."*

Though the remaining three insulted teachers began the year skeptical of INVEST, it was not until they did not reach Level 4 that frustration materialized. One of these teachers shared, *"I thought it was a good program at the beginning…. But to me now, I see it's very unfair. I thought I was doing everything that was highly effective. I didn't change anything I've done in the last year when I was always such a strong teacher….So did I get my kids where they needed to be? Did I do everything I was supposed to do as a teacher? I went above and beyond"* (School 4, Teacher 1). These teachers were very conscientious and had very high expectations for their performance. When they did not reach Level 4, they took the decision personally and reacted in an emotional manner.

Skeptical Teacher Profile

Overall, the most common category was the skeptical teacher. Unlike other groups, skeptical teachers did not have strong reactions to INVEST and remained fairly neutral toward the system over the course of the year. The majority of the skeptical teachers appeared to be well-established and respected at their respective schools and had performed well on the new system. At the beginning of the year, they were initially skeptical and worried about specific details of INVEST, but over the course of the year, the system did not appear to have much influence on their practice or attitudes toward their work. With the exception of one first year teacher (who had done her student teaching in Aldine), all 13 of the skeptical teachers had more than five years of experience in the classroom and seemed well versed to changes in district policy. Every

school had at least one skeptical teacher and it was the dominant type on several campuses (Schools 4 and 5 in particular).

Skeptical teachers had similarities to other personality profiles but also some unique differences. Like the majority of teachers, they reported being conscientious and generally agreeable in their approach to their work with colleagues. As was the case with invested veterans, they were quite confident in their ability to meet students' needs and believed they controlled their practice. However, as distinct from other types, skeptical teachers appeared to be highly analytical, practical individuals. They raised very specific questions associated with the system's implementation and seemed to have a more realistic attitude as to what type of changes would actually happen as a result of INVEST. In some cases, this analytic nature bordered on neuroticism (or a lack of emotional stability), but for the most part, these teachers felt they were being practical about what types of proposed changes were feasible to implement in practice.

As the year progressed, several skeptical teachers became more open to implementing changes suggested by INVEST. One teacher shared, *"I adjust to whatever makes sense,"* while another commented, *"you can always see ways to grow."* Unlike insulted teachers, they were not frustrated by not meeting the system's expectations as long as they were meeting their own expectations, which they viewed as paramount. A subset of skeptical teachers remained apathetic over the course of the year and unaffected by any of the system's suggestions. As one of these teachers shared, he was unfazed when he did not receive the highest score on INVEST, because the important part was that he met his own expectations: *"I won't say I don't care, but I don't know, if I'm*

proficient, OK. I think that as long as I know I'm doing what I'm supposed to do in the classroom, I'm OK" (School 5, Teacher 5).

Section Two: Key Individual Characteristics

As these profiles make clear, teachers' personality and subsequent reactions to INVEST varied considerably across the sample of teachers I interviewed. In this section, I used multiple regression analyses to examine which personality characteristics were most influential in predicting teachers' motivation, effectiveness, and retention on the new system in the population of teachers and then revisited these teacher profiles to help explicate findings. As discussed in Chapter 2 (and presented in Table 2-3), I used the following personality characteristics listed below (each of these measures is on a scale of 1-5 with 1 being strongly disagree and 5 being strongly agree):

- ***Teaching grit*** – teachers' passion and perseverance for their teaching. The survey captured information on teachers' interest in teaching, whether setbacks have discouraged them and how important it was to them to improve their teaching.
- ***Overall grit*** – teachers' passion and perseverance more generally. The survey captured information on whether teachers were diligent, stayed focused on goals, and finished whatever they began.
- ***Conscientiousness*** – teachers' desire to do a task well. The survey captured information on whether teachers were thorough, efficient, and followed through on their obligations.

- ***Extraversion*** – teachers' energy and level of engagement with others. The survey captured information on how talkative, outgoing, and sociable teachers are as individuals.
- ***Agreeableness*** – the value teachers place on getting along with others. The survey captured information on how forgiving, considerate, and kind teachers were as individuals.
- ***Emotional Stability (opposite of Neuroticism)*** – teachers' degree of impulse control and lack of anxiety. The survey captured information on how relaxed teachers were and how well they handled stress.
- ***Openness*** – teachers' appreciation for new ideas and level of curiosity. The survey captured information on how original, active, and open teachers were to new experiences.

As we would expect from prior research, these personality characteristics were all positively correlated with each other, though the magnitude of the correlations were relatively small in size. Table 5-2 summarizes the correlations between the various personality characteristics. Interestingly, grit and grit in teaching were only moderately correlated ($r = .32$, $p < .001$), suggesting that grit is somewhat domain-specific, and there are challenges that are unique to teaching.

Table 5-2

Correlations between Teachers' Individual Characteristics

Measures (Scale 1-5)	Mean (SD)	1	2	3	4	5	6	7
1. Teaching Grit	3.80 (0.67)	-	0.32***	0.27**	0.04	0.21***	0.15***	0.17***
2. Overall Grit	3.90 (0.48)		-	0.58***	0.06**	0.30***	0.29***	0.24***
3. Conscientious-ness	4.33 (0.50)			-	0.09***	0.32***	0.16***	0.28***
4. Extraversion	3.39 (0.80)				-	0.01	0.04	0.24***
5. Agreeableness	4.20 (0.58)					-	0.25***	0.21***
6. Emotional Stability	2.37 (0.87)						-	0.08***
7. Openness	4.12 (0.55)							-

**p<.05. **p<.01. ***p<.001*

Teachers' personality characteristics were also associated with their motivation and performance on the new system. As demonstrated in Table 5-3, these individual characteristics tended to be more correlated with system motivation (expectancy and value), reported changes in practice, and burnout than with performance or turnover from the district. Of the personality characteristics, teaching grit was the most positively correlated with teachers' belief in their ability to improve practice on the new system (system expectancy), $r = .29$, $p < .001$, the value they placed on the new system (system value), $r = .28$, $p < .001$, and their reported changes in practice, $r = .29$, $p < .001$, as well as the most negatively correlated with their level of burnout, $r = -.53$, $p < .001$, and turnover from the district, $r = -.10$, $p < .05$. In other words, teachers with more long-term passion and perseverance for their teaching were more likely to be motivated on

INVEST, to report implementing changes in their practice, and to avoid becoming burnt out by their work. Though teaching grit was positively correlated with performance on the Danielson Framework (r = .09, p < .01), other personality characteristics had slightly stronger associations. In particular, teachers' conscientiousness was positively correlated with teachers' performance on the Danielson Framework, r = .16, p < .001. Conversely, teachers' level of emotional stability was negatively associated with their performance on the Danielson Framework, r = -.11, p < .001. In other words, teachers who were very

Table 5-3

Correlation between Individual Characteristics and Teacher Outcomes

Measure	System Expectancy	System Value	Danielson Framework	Student Growth Percentiles	Reported Changes in Practice	Turnover	Burnout
Individual Characteristics							
Teaching Grit	0.29***	0.28***	0.09**	0.02	0.29***	-0.10*	-0.53***
Overall Grit	0.11***	0.17***	0.08*	0.02	0.10**	-0.04	-0.30***
Conscientiousness	0.09**	0.19***	0.16***	-0.02	0.03	-0.00	-0.21***
Extraversion	0.05	0.07*	-0.01	0.03	0.06*	0.00	0.00
Agreeableness	0.09**	0.14***	-0.07*	-0.04	0.14***	-0.01	-0.13***
Emotional Stability	0.09**	0.05	-0.11***	0.14*	0.02	-0.02	-0.27***
Openness	0.13***	0.12***	0.00	-0.07	0.10**	0.02	-0.05
School Characteristics							
Quality of administration	0.30***	0.25***	0.12***	0.04	0.29***	-0.14***	-0.39***
Positive support	0.26***	0.18***	0.11***	0.02	0.26***	-0.10**	-0.38***
Level of control	0.23***	0.22***	0.12***	0.08*	0.26***	-0.09**	-0.35***
Professional community	0.18***	0.15***	-0.04	0.02	0.27***	-0.05	-0.27***
System Characteristics							
Impact Factor	0.47***	0.33***	0.01	-0.04	0.52***	-0.11***	-0.33***
Observation Factor	0.44***	0.37***	0.09**	0.01	0.45***	-0.06	-0.24***
Understanding Factor	0.40***	0.41***	0.08*	0.02	0.46***	-0.09**	-0.27***
Support Factor	0.45***	0.35***	0.01	-0.02	0.55***	-0.10**	-0.25***
Student Growth Factor	0.50***	0.34***	0.05	-0.01	0.44***	-0.03	-0.31***
Goal Factor	0.38***	0.36***	0.03	-0.02	0.54***	-0.09**	-0.25***

Note: For all measures, N=1097, except for Student Growth Percentiles measure, where N=651. *$p < .05$. **$p < .01$. ***$p < .001$

diligent performed better on the new system, and teachers who were neurotic and anxious about the new system also appeared to have been driven to higher levels of performance.

Individual Characteristics and Teacher Motivation

This section explores how the first outcome of interest – teacher motivation – was influenced by these individual characteristics. Recall that motivational choices are linked to two distinct (but related) sets of beliefs: an individual's expectancy (or belief in their ability) and an individual's value (or the importance associated with a particular task). Both are necessary conditions for individual behavioral change. In the case of teachers, if they doubt their ability to reach certain standards (lower expectancy) or do not find it important to meet those expectations (lower value), they will not be motivated to improve their performance. As discussed in Chapter 4, overall motivation for teaching is associated, but not synonymous, with motivation to perform well on the new evaluation system. Since this chapter is concerned with the implementation of INVEST, it will primarily focus on teachers' expectancy and value on the system itself.

To see which of the personality characteristics predicted teacher motivation above and beyond other factors, I ran multiple regression analyses where I simultaneously entered all of the individual characteristics into the model. Then, to assess the predictive power of individual characteristics, I entered two new sets of predictors (i.e., school characteristics, system characteristics) sequentially to determine the influence of individual characteristics when controlling for school and system factors. To account for other baseline differences, I controlled for individual-level characteristics which were significant predictors in the difference-in-differences modeling (i.e., ethnicity,

Table 5-4

Regression Analysis Predicting System (INVEST) Expectancy

Measure	Without School Effects			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.28(0.03)***	0.17(0.04)***	0.08(0.04)*	0.25(0.04)***	0.15(0.04)***	0.08(0.04)
Overall Grit	-0.02(0.04)	-0.03(0.04)	-0.02(0.04)	-0.03(0.04)	-0.04(0.04)	-0.02(0.03)
Conscientiousness	-0.05(0.04)	-0.03(0.05)	0.00(0.04)	-0.05(0.04)	-0.03(0.04)	-0.01(0.04)
Extraversion	0.03(0.03)	0.04(0.03)	0.04(0.03)	0.01(0.03)	0.02(0.03)	0.03(0.03)
Agreeableness	0.01(0.04)	-0.01(0.04)	-0.01(0.03)	0.02(0.04)	-0.00(0.04)	-0.01(0.04)
Emotional Stability	0.00(0.03)	0.00(0.03)	-0.03(0.03)	-0.00(0.03)	-0.00(0.03)	0.04(0.03)
Openness	0.09(0.04)*	0.09(0.04)*	0.09(0.04)*	0.10(0.04)**	0.09(0.04)*	0.08(0.04)*
School Characteristics						
Quality of Administration		0.16(0.05)**	0.08(0.05)		0.14(0.05)**	0.08(0.05)
Positive Support		0.09(0.06)	0.03(0.05)		0.11(0.05)*	0.02(0.05)
Level of Control		0.04(0.04)	-0.01(0.04)		0.06(0.04)	-0.00(0.04)
Professional Community		-0.03(0.04)	-0.07(0.04)		-0.02(0.04)	-0.05(0.04)
System Characteristics						
Impact Factor			0.05(0.09)			0.03(0.09)
Observation Factor			0.12(0.06)*			0.08(0.06)
Understanding Factor			-0.14(0.07)			-0.09(0.07)
Support Factor			0.28(0.09)**			0.25(0.09)**
Student Growth Factor			0.37(0.10)***			0.36(0.09)***
Goal Factor			-0.10(0.09)			-0.07(0.10)

Note. All continuous variables have been standardized. This analysis was conducted on teachers in the pilot schools who had completed both the beginning and end of year survey (N = 1097). It controls for all individual-level demographic characteristics.
*p < .05. **p < .01. ***p < .001

certification, and years of teaching) and conducted my analysis with and without school fixed effects. Tables 5-4 and 5-5 display the results from multiple regression analyses used to examine which factors predicted teachers' system expectancy (or belief in their ability to perform well on INVEST) and teachers' system value (or how important it was for them to perform well on INVEST). In this section, I discuss which individual

characteristics were most influential in predicting teacher motivation on the new system and use the qualitative profiles to explicate results.

Table 5-5

Regression Analysis Predicting System (INVEST) Value

Measure	Without School Effects			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.26(0.04)***	0.17(0.05)**	0.09(0.04)*	0.24(0.04)***	0.15(0.04)**	0.09(0.04)*
Overall Grit	0.03(0.04)	0.03(0.04)	0.03(0.04)	0.03(0.04)	0.03(0.05)	0.03(0.04)
Conscientious-Ness	0.08(0.04)*	0.09(0.04)*	0.09(0.04)*	0.08(0.04)	0.08(0.04)*	0.08(0.04)*
Extraversion	0.05(0.03)	0.06(0.02)	0.04(0.03)	0.04(0.03)	0.05(0.03)	0.05(0.03)
Agreeableness	0.06(0.04)	0.04(0.04)	0.04(0.04)	0.05(0.04)	0.03(0.04)	0.04(0.04)
Emotional Stability	0.04(0.04)	-0.04(0.03)	-0.03(0.03)	-0.02(0.04)	-0.01(0.04)	-0.01(0.03)
Openness	0.02(0.04)	-0.00(0.04)	0.01(0.04)	0.04(0.04)	0.01(0.04)	0.02(0.04)
School Characteristics						
Quality of Administration		0.12(0.05)*	0.06(0.04)		0.15(0.05)**	0.08(0.05)
Positive Support		0.04(0.04)	-0.02(0.04)		0.05(0.05)	-0.05(0.04)
Level of Control		0.05(0.03)	0.02(0.03)		0.07(0.03)	0.02(0.03)
Professional Community		0.01(0.04)	-0.01(0.04)		-0.01(0.05)	-0.04(0.05)
System Characteristics						
Impact Factor			-0.04(0.09)			-0.00(0.09)
Observation Factor			0.12(0.06)*			0.13(0.05)*
Understanding Factor			0.21(0.07)**			0.22(0.08)**
Support Factor			0.07(0.08)			0.08(0.09)
Student Growth Factor			0.01(0.10)			0.02(0.11)
Goal Factor			0.10(0.07)			0.09(0.07)

Note: All continuous variables have been standardized.
*$p < .05$. **$p < .01$. ***$p < .001$.

Grit. Teachers' system motivation – both expectancy and value – was influenced by their level of teaching grit. In Table 5-4, Model 1 showed that grit in teaching significantly influenced teachers' level of system expectancy, $B = .28$, $t(908) = 9.78$, $p < .001$, and Model 4 revealed that this was robust to the inclusion of school effects. In

addition to believing in their ability to perform well on INVEST, grittier teachers were also more likely to value their performance. Table 5-4 Model 1 confirmed that grit in teaching was a significant predictor of system value, $B = .26$, $t(908) = 10.58$, $p < .001$ and again, Model 4 validated that this effect was still significant with school effects.

In the qualitative data, invested teachers (the profile type which most exemplified grit for teaching) valued INVEST and took ownership over ensuring their classrooms were student directed. When they received a low observation score, they looked internally to their practice, rather than blaming the system. One invested teacher explained how she used observation data to reflect on her practice, "*Did I not teach you this? Did you not understand? And that's what I use to improve myself. What did I do wrong that they did not get"* (School 1, Teacher 6). To the contrary, burnt-out teachers questioned their ability to reach a more challenging student population, which led them to give up when faced with obstacles. As one teacher shared, *"students have mentally checked out. I don't care what you say, they're 15 year old boys and girls. Sorry. It's not a perfect world"* (School 6, Teacher 3). In part because of this lack of grit, performing well on the system did not seem to be a realistic possibility and resulted in lower system expectancy.

Conscientiousness. Teachers' system value was also influenced by their level of conscientiousness. Model 1 in Table 5-4 established that conscientiousness was a significant predictor, $B = .08$, $t(908) = 10.58$, $p < .05$, and again Model 4 demonstrated this was still significant with school effects in the model. Across all personality types, teachers appeared to be organized, dependable, and appreciated structure. Consequently,

this was one of the reasons they were drawn to the clear and systematic nature of the Danielson Framework, as it provided them with a thorough understanding of the expectations. Many teachers displayed an almost dutiful nature when engaging in work-related tasks. As one teacher shared, *"For me, this is where the bar was set. I've got to learn to work with this. This is my baseline here"* (School 4, Teacher 3).

Openness. Despite being consistently conscientious, there appeared to be considerably more variation in teachers' openness across types which influenced system expectancy. One skeptical teacher shared, *"I've been teaching 15 years, at the beginning of the year, I'm thinking what's wrong with PDAS. I like PDAS and all of a sudden they just changed it"* (School 2, Teacher 3). Skeptical veterans were traditional in their outlook and tended to be closed off to new experiences. They may have begun their teaching more optimistic about new system's possibilities (as was the case with novices) but given challenges over the years, their system expectancies had lowered over time. Insulted teachers (like the one below) felt blamed for challenges, which led to lower openness to new policies:

> It's a pie in the sky theory which is lovely but it's impractical…it seems like teachers are always being told all the time, we've got kids who never show up to school because we're not building relationships with them. We've got kids dropping out of school, because we're not offering the proper interventions. There is only so much we can do and now they're actually going to ding us for that? What the hell. Not this again (School 6, Teacher 2).

Model 1 in Table 5-3 showed that teachers' openness to new experiences was a significant predictor, $B = .09$, $t(908) = 9.78$, $p < .05$, of teachers' system expectancy (or their belief in their own abilities as educators) and this effect remained consistent when controlling for school effects.

Individual Characteristics and Teacher Effectiveness

Theory indicates that even if teachers are motivated to improve performance, this will not necessarily lead to changes in teachers' effectiveness. Indeed, initial motivation must be translated into targeted action and that action must be sustained over an extended period of time. When individuals move from the deliberative to the action phase, they make a commitment to a specific set of next steps and engage in self-regulatory planning. This section examines how teachers engaged in this improvement process and how the second outcome of interest – teacher effectiveness – was influenced by individual characteristics.

As discussed in Chapter 4, there were limitations to the measures used to capture teachers' effectiveness. For volition, I used an end of year survey question which asked teachers to indicate whether they had implemented changes in their practice as a result of the new system. Teachers' effectiveness was captured through two measures – (1) Danielson Framework, which was the final average of the observation scores throughout the pilot year and (2) Student Growth Percentiles, which assessed how much teachers' students grew in relation to other students beginning the year at a similar starting point. SGPs were only available for teachers in tested subjects (in pilot schools $N = 302$) compared to the Danielson measure ($N = 1097$). Additionally, though teachers were familiar with the SGP measure, they did not receive their results during the pilot. Since this chapter is concerned with how teachers responded to INVEST, the Danielson Framework is the better for that purpose. To see which of the personality characteristics predicted teacher effectiveness above and beyond other factors, I ran multiple regression

analyses using the same methods discussed above. Table 5-6 presents the results examining which factors predicted teachers' effectiveness (as measured by the Danielson Framework for Teaching).

Table 5-6

Regression Analysis Predicting Teacher Effectiveness on Danielson Observation Measure

Measure	**Without School Effects**			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.09(0.04)*	0.03(0.05)	0.03(0.04)	0.12(0.04)**	0.07(0.05)	0.06(0.04)
Overall Grit	0.01(0.05)	0.00(0.04)	0.01(0.04)	0.00(0.04)	-0.01(0.04)	-0.00(0.04)
Conscientious-ness	0.15(0.04)***	0.15(0.04)**	0.14(0.04)***	0.15(0.04)***	0.14(0.04)***	0.13(0.04)**
Extraversion	-0.02(0.04)	-0.02(0.03)	-0.02(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)
Agreeableness	-0.10(0.03)**	-0.10(0.03)**	-0.10(0.03)**	-0.10(0.03)**	-0.09(0.03)**	-0.09(0.03)**
Emotional Stability	-0.12(0.03)**	0.12(0.03)***	0.11(0.03)***	-0.09(0.03)**	-0.09(0.03)**	-0.09(0.03)**
Openness	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.02(0.03)
School Characteristics						
Quality of Administration		0.11(0.04)*	0.08(0.05)		0.12(0.04)**	0.09(0.04)*
Positive Support		0.07(0.03)*	0.08(0.04)*		0.07(0.03)	0.07(0.04)
Level of Control		0.09(0.03)**	0.07(0.03)*		0.10(0.03)**	0.08(0.03)*
Professional Community		-0.14(0.04)***	-0.12(0.04)**		-0.19(0.04)***	-0.16(0.05)**
System Characteristics						
Impact Factor			-0.07(0.08)			-0.07(0.08)
Observation Factor			0.11(0.06)*			0.09(0.06)*
Understanding Factor			0.13(0.06)*			0.18(0.07)*
Support Factor			-0.16(0.10)*			-0.15(0.09)*
Student Growth Factor			-0.03(0.08)			-0.02(0.08)
Goal Factor			0.03(0.06)			-0.01(0.07)

Note: All continuous variables have been standardized.
*$p < .05$. **$p < .01$. ***$p < .001$

Results-Driven (Conscientiousness, Grit, and Emotional Stability). As demonstrated in Model 1 in Table 5-6, conscientiousness and teaching grit both emerged

as significant predictors of teachers' scores on the Danielson Framework, and Model 4 indicated that these results were robust to the inclusion of school effects. In Model 1, conscientiousness had the largest coefficient ($B = .15$, $t(890) = 7.87$, $p < .001$) followed by grit in teaching ($B = .09$, $p < .05$). This data was confirmed in the qualitative data, as teachers who reported being driven by results (invested teachers) were more likely to reach higher levels of performance under the new system. They set specific goals and filtered feedback they received from their evaluators through this lens. Further, invested teachers used feedback to track progress on their goals; as one teacher shared: *"it's great because it gives me a very clear picture from this point forward of what I need to do differently"* (School 1, Teacher 3). When they were at schools where administrator feedback was limited, invested teachers used the rubric themselves to benchmark progress and drive improvements. *"Myself, I'm driven by results. I'll say the data drives me. I'm always asking what could I have done differently and INVEST helps me do that reflection"* (School 1, Teacher 6).

Sometimes teachers' obsession with goals bordered on neuroticism and led to increased stress and high levels of anxiety (or lower emotional stability). This level of ambition made coping with new expectations (which were rolling out over the course of the year) quite stressful:

> It's too much, don't add on it. If we are adding on, I'm getting all confused. If it's a pilot, let it run. What happens, happens. We'll change it next year. Add on in the beginning. You said do this. We'll do it and see how it goes. And after that, next year add on whatever. We have to get used to this…I keep going am I asking do I do this. Am I doing this. So at this time I'm overwhelmed but I'm loving it.

Though this level of commitment would be hard to sustain over time (see Retention section below), it appeared to help drive improvements in teacher practice (at least in the short-term), as demonstrated in Table 5-6.

Individual Characteristics and Teacher Retention

In the aggregate, when asked whether INVEST would influence their retention, most teachers did not report that it had had much influence over their individual decision. However, many teachers (particularly burnt-out and insulted teachers) worried about the long-term impact the new system would have on the district's ability to retain talented educators. Their concerns were primarily rooted in the system's impact on teacher burnout. This section explores which individual characteristics were associated with teachers' level of burnout and discusses how the third outcome of interest – teacher retention – may be influenced as a result.

As demonstrated in Table 5-3 above, specific personality characteristics were associated with teachers' level of burnout and turnover. To assess the incremental predictive validity of these individual characteristics, I ran a binary logistic regression model to examine which factors predicted teachers' retention while controlling for other school and system factors. For this analysis, I compared the stayers with the leavers to be consistent with the analysis conducted in Chapter 5. Additionally, because INVEST was being implemented in all schools the following year, I was more interested in whether teachers had left the district entirely, rather than whether they moved to another school within the district. I conducted these analyses using the same methods discussed for teacher motivation.

Table 5-7

Binary Logistic Regression Table Predicting Turnover

Measure	Without School Effects			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.79(0.05)**	0.88(0.06)	0.83(0.10)	0.79(0.06)**	0.85(0.07)*	0.82(0.11)
Overall Grit	0.92(0.08)	0.94(0.08)	0.94(0.11)	0.92(0.08)	0.93(0.08)	0.91(0.13)
Conscientious-ness	1.18(0.16)	1.17(0.10)	1.19(0.17)	1.18(0.12)	1.17(0.11)	1.25(0.19)
Extraversion	0.92(0.06)	0.92(0.05)	0.99(0.11)	0.90(0.05)	0.90(0.05)	0.99(0.10)
Agreeableness	1.01(0.08)	1.04(0.08)	1.11(0.12)	1.01(0.08)	1.03(0.08)	1.10(0.13)
Emotional Stability	1.04(0.09)	1.05(0.07)	*0.96(0.12)*	*1.09(0.08)*	1.10(0.12)	1.01(0.13)
Openness	1.04(0.09)	1.04(0.09)	1.09(0.15)	1.05(0.09)	1.05(0.09)	1.11(0.16)
School Characteristics						
Quality of Administration		0.82(0.06)**	0.68(0.07)**		0.81(0.06)**	0.70(0.08)**
Positive Support		0.98(0.08)	0.97(0.10)		0.94(0.08)	0.86(0.09)
Level of Control		0.89(0.06)	0.91(0.12)		0.90(0.06)	1.00(0.15)
Professional Community		1.04(0.07)	1.11(0.12)		1.15(0.08)	1.22(0.32)
System Characteristics						
Impact Factor			0.41(0.15)**			0.40(0.14)*
Observation Factor			0.98(0.15)			1.02(0.17)
Understanding Factor			0.85 (0.19)			0.87(0.20)
Support Factor			1.32(0.32)			1.22(0.32)
Student Growth Factor			3.13(1.07)**			3.38(1.30)**
Goal Factor			0.92(0.26)			0.93(0.27)

Note: All continuous variables have been standardized. Only 873 teachers had data on all variables included in each analysis, so this was the final analytic sample used for all models. Teachers who left the district were slightly less likely to complete end of year surveys, which explains why the aggregate turnover rates are lower than those reported in Chapter 4. *$p < .05$. **$p < .01$. ***$p < .001$

Grit. This analysis confirmed previous research that gritty individuals work diligently and avoid burnout when confronting obstacles or setbacks in performance. As demonstrated in Model 1 in Table 5-7, teaching grit was the only personality variable to emerge as a significant predictor of teachers' retention. Teachers who were one standard

deviation higher in grit in teaching were 21% less likely to leave teaching in the district, *OR* = .79, $p < .01$. However this coefficient was not significant when controlling for school and system characteristics. No other personality variables were significant predictors of teacher retention. Indeed, stayers had significantly higher grit in teaching, *M* = 3.79, *SD* =.68, when compared to Leavers *M* = 3.57, *SD* = .78, $p < .01$.

In the qualitative data, invested teachers (who reported intending to stay in teaching for the long-term) remained positive over the course of the year and were not as overwhelmed as other personality types by the additional workload. Instead of viewing the paperwork as a nuisance, they found ways to use the process to drive their teaching. One invested teacher remarked that *"If this is for helping me as a teacher, I love it. If I'm not proficient or excellent, help me. Tell me how to do it. Help me be a better teacher. I think they are trying to find out more effective to teachers to make more effective students"* (School 1, Teacher 1). As high achievers, they believed they were never perfect and always wanted to do more to push their practice to the next level. One teacher shared, "*In order for you to grow, you have to be able to accept constructive criticism and see how to use it to make you better*" (School 1, Teacher 3). Not all teachers were as open to feedback and as a result, many became very discouraged when they did not reach Level 4. As discussed in Chapter 4, this lowered expectancy resulted in a higher degree of burnout among a subgroup of teachers.

Emotional Stability. Across all personality types, teachers were very conscientious individuals, which made them (on the whole) eager to meet expectations. However, when this level of conscientiousness was taken to an extreme (and was not

accompanied by the gritty mindset discussed above), it caused anxiety among educators about the purpose of new INVEST. Many of the skeptical and insulted veterans had lower levels of emotional stability and believed the system was intended to be a "weapon" or "tool used to punish teachers." Several of these teachers had read national news about efforts in Washington, DC and Chicago to fire ineffective teachers and believed INVEST was part of a national conspiracy to "blame teachers" for problems of poverty. Other neurotic teachers assumed the system was merely a way for the district to get more funds from the state and that it had been mandated, rather than locally developed. As the teacher's comment below suggests, these subsets of teachers were not buying the purpose as it had been communicated to them by their principal. "*They want me to think INVEST is about assessing teachers to see whether or not they're doing what they need to do to reach the children. That's what I think I'm supposed to think. What I think is though, it's all about money*" (School 5, Teacher 5). Regardless of how they explained the system to their teachers, every principal shared that they had a few who remained very skeptical of their intentions.

Summary

In sum, this chapter demonstrated that teachers' responses to INVEST varied considerably and could be categorized into one of five distinct personality profiles – invested teachers, sponge teachers, burnt-out teachers, insulted teachers, and skeptical teachers. Invested teachers were veteran teachers who were open to new experiences, displayed grit by engaging in strategic goal-setting, and remained conscientious in working towards desired outcomes. Sponge teachers were novices who were very open

and conscientious which was driven by their agreeable nature and strong desire to prove themselves as new teachers. Though they were primarily new teachers, burnt-out teachers were less open and agreeable than their sponge counterparts and lacked the grit necessary to stay focused on long-term goals. Insulted teachers reported being very focused on achieving highly effective status on the new evaluation system, and as a result, when they did not reach high standards, they were incredibly frustrated and negative about the system. Skeptical teachers were practical and analytical individuals, which made them initially less emotionally stable in response to change. Some skeptical veterans were entirely apathetic while others were just less likely to embrace reform until they had seen how it played out in practice.

Of all of the personality characteristics, one emerged as predictive across all key outcomes of interest – teaching grit. In comparison to other types, the invested teachers appeared most likely to exemplify gritty traits of passion and perseverance for long-term goals. Though other types began the year open to the possibility of change, invested teachers remained committed to the pursuit of enhancing their teaching over time. They used the Danielson Framework very deliberately to identify specific areas for improvement and aligned long-term goals with corresponding plans of action. When faced with challenge, they internalized the need for change and were able to incorporate feedback into their practice. Rather than being burnt-out by the additional workload, invested teachers identified how the new expectations reinforced their personalized goals for improving performance and worked strategically to implement changes in a meaningful way. Though teachers' gritty nature seemed to be at least somewhat internal,

it does appear that their motivation and subsequent performance was also influenced by contextual features. Indeed, three of the five invested teachers taught in the same school, and several campuses did not have any teachers falling into this type. While this could be by chance, it seems likely that context also influenced teachers' approaches to their work and responses to INVEST. The next chapter will explore these organizational school-based factors in more detail.

CHAPTER 6: SCHOOL-LEVEL VARIATION

Research has demonstrated that teachers' motivation, performance, and retention can be influenced by school-level variables. In particular, prior studies have investigated how professional community, the quality of principal leadership, and the level of teacher involvement in decision-making structures influence teachers' motivation (Kelley et al., 2000; Rosenholtz, 1989). In order to successfully improve practice, teachers need consistent feedback on their performance. Given the design of new evaluation systems, the principal is essential in providing this type of support and driving professional learning. Moreover, teachers' satisfaction and subsequent decision to remain in the profession has been shown to be positively associated with measures of autonomy and faculty influence (Ingersoll, 2001; Ingersoll, 2006). In this section, I explore how variation in these school-level factors predicted outcomes of interest by presenting a case study for each of the six schools I visited (Section One) and then examining which school characteristics were most predictive of outcomes under INVEST (Section Two).

- ***Section One: School Profiles***. Though my teacher interviews indicated variation at the individual level (within schools), I also noticed trends across campuses (between schools). Indeed, specific profile types of teachers appeared more frequently on some campuses than others. At certain schools, teachers believed INVEST would drive improvements, while teachers at other schools worried the system would drive good teachers out of the district. The principals I interviewed also expressed varying degrees of comfort with the new system, which subsequently influenced varied implementation across schools. As discussed in

Chapter 2, I developed codes in Atlas.ti to help interpret these divergent responses and created school profiles for each campus I visited.

- ***Section Two: Key School Characteristics.*** After examining these school profiles, I then set out to systematically investigate how school climate influenced outcomes within the larger population of teachers in the district. Although many school climate surveys exist, I chose to use items from the National Center for Education Statistics (2012) teacher questionnaire which has been vetted and used with a nationally representative population of teachers. In this section, I examine how teachers' perceptions of school climate influenced teachers' motivation, effectiveness, and retention.

Section One: School Profiles

This section describes overall trends in attitudes and perceptions for each of the six case study schools. Table 6-1 provides a short description of each type and a summary of their reactions to INVEST. In the Appendix, I provide more information on the school-level survey data. After exploring this qualitative data, I use these profiles, as well as the survey data, to examine how key climate indicators predict teacher outcomes – i.e., motivation, effectiveness, and retention.

Table 6-1

School Profiles

School Profiles	Description	Overall Reactions to INVEST
School 1	School 1 is an ***underperforming*** elementary school based on the state rating system and according	At the beginning of the year, the School 1principal was overwhelmed by the increased

	to the principal and veteran teachers. School 1 serves a high need primarily African-American student population. Despite challenges, School 1 ***has a positive and collaborative culture*** and teachers place strong ***trust in the administration***. The principal gives teachers considerable ***control and autonomy*** over their practice.	workload associated with the implementation of INVEST but had developed new structures by the end of the year. The principal used the Danielson Framework to provide detailed feedback and empower teachers to drive their own improvement process. Overall, teachers had a positive attitude towards INVEST. However, many teachers were overwhelmed by the increased paperwork associated with the new system.
School 2	School 2 is consistently one of ***the highest performing*** elementary schools in the district and serves a primarily Hispanic student population. Generally, the ***principal is highly respected*** by her teachers, though some teachers expressed concerns about a cliquey nature at the end of the year. The principal has ***high expectations for performance*** and the school has a clear culture of achievement.	INVEST was implemented with fidelity at School 2, with the exception of post-observation conferences which appeared to be less frequent. Despite successful implementation, teachers at School 2 had mixed attitudes towards INVEST. While many appreciated the detailed and focused feedback, others were frustrated by the unattainability of Level 4 performance on the Danielson Framework for Teaching. In particular, teachers expressed concern over the accuracy of the observation results.
School 3	School 3 is an intermediate school that was underperforming prior to INVEST implementation but was identified as meeting standards during the pilot year. The ***principal is respected*** as a leader and the school has a ***strong culture of collaboration***. The principal uses staff development time to discuss new initiatives which helps ***build teachers' understanding and***	The principal and teachers at School 3 generally had quite positive attitudes towards INVEST. The principal was extremely organized and able to scaffold the introduction of the new system for teachers to build understanding and investment. Teachers at School 3 did not appear to be as concerned with the additional workload of INVEST and most teachers felt supported

	trust in new systems.	rather than intimidated by the observation process.
School 4	School 4 is one of the highest performing intermediate schools in the district and serves a primarily Hispanic population. The school has a ***positive climate***, which in large part is due to the ***principal's strong reputation*** among his teachers. School 4 has a ***significant percentage of veteran teachers*** who have been teaching at the school for a number of years.	Though there were some challenges associated with an increased student enrollment, INVEST was generally implemented with fidelity. Yet teachers had mixed attitudes towards the new system. Though they felt empowered by the pre-conferences and appreciated the additional observations, many of the veterans were frustrated by not attaining the highest level of performance. Novices were generally more receptive to the feedback offered by the new system.
School 5	School 5 is typically an underperforming ninth grade school, but met expectations during the pilot year for the first time in many years. School 5 is half African-American/half Hispanic and according to the principal and teachers, serves a challenging student population. The ***principal is well-respected*** by the staff, but the school has a reputation for being d***isorganized*** and having a more ***skeptical community of veteran teachers.***	INVEST was not implemented with fidelity at School 5 and at the end of the year, many teachers had yet to receive feedback on their performance. Teachers appreciated the opportunity to reflect using INVEST but felt they were working harder without being recognized. In particular, teachers raised concerns over the inappropriateness of the Level 4 expectations on Danielson (student-centered classrooms) and the unfairness of Student Growth Percentiles (given the lack of student accountability) at the high school level.
School 6	School 6 is a ninth grade school which ***has historically been one of the flagship schools*** in the district but for the first time in many years, it did not meet	School 6 struggled to implement INVEST with fidelity, in part because of the significant percentage of new teachers on their campus. The principal

standards. School 6 has a very structured environment with a distinct culture. Some teachers appreciate the ***orderly environment***, while others are frustrated by what they perceive to be ***unrealistic expectations imposed by the administration***. School 6 has a consistently high turnover rate with a ***significant percentage of new teachers each year.***	believed that the new system resulted in constructive feedback and useful dialogue; however, teachers' perceptions were more mixed. Most teachers reported that they appreciated receiving feedback, but the feedback was not consistently provided. Several teachers viewed the new system as additional – and unnecessary work – and reported that it led to lower morale among the teachers.

School 1 Overview: Lower Performing Elementary School

According to the principal and several of the veteran teachers, School 1 serves a high need student population. Despite challenges, School 1 had a positive and collaborative school climate, the principal was held in high regard and teachers generally felt supported to improve their practice. Additionally, teachers reported they had the flexibility and autonomy to make decisions in their classrooms that met their students' needs, which built strong trust with the administration. According to one teacher, it is the freedom that keeps her teaching at School 1: *"I like the freedom because I know what works so I can best meet the needs of my students."*

On average, teachers at School 1 had 10 years of experience, with 25% of the faculty in their first five years of the profession. Given the small size of the school, the teachers had a strong professional community, which met frequently (weekly) to discuss new initiatives. The six teachers I interviewed at School 1 all had prior teaching experience but two were new to the school. Three of the six teachers were invested veterans and remained motivated over the course of the year to use the feedback to

improve their teaching. Two of the six teachers fell into the skeptical subgroup; in their minds, INVEST was "*something you have to get done just like anything else*." The final teacher was a burnt out novice – though she had prior teaching experience, she was incredibly overwhelmed and frustrated by what she considered to be unreasonable expectations in her first year teaching in Aldine.

The majority of teachers saw the purpose of INVEST as improving teachers' performance and giving them the tools they needed to succeed. At the beginning of the year, one teacher shared: "*I think it's trying to help teachers. I honestly believe that."* When INVEST was presented in initial trainings, the principal focused on how INVEST was an improvement on PDAS (Professional Development and Appraisal System) and used time in the school's strong professional learning community to view the Teachscape modules together in groups and discuss expectations. As a result of this framing, most teachers welcomed INVEST as a tool to guide their development. One teacher shares that "*unlike PDAS, INVEST is about providing a straightforward review for teachers throughout the year and helping them become better."*

This perspective on INVEST stayed fairly consistent over the course of the year, and teachers still felt positive about the systems' intentions at the end of the year. Though the principal struggled with implementation in the fall, she had developed new systems for staying on track with the process by the end of the year. The principal used the Danielson Framework to provide specific and detailed feedback on areas to improve practice and teachers were empowered to drive their own reflection process. However, though teachers appreciated the comprehensive nature of the evaluation system, this also

made it very overwhelming. Coupled with what many teachers viewed to be *"excessive and unnecessary paperwork,"* INVEST had increased teachers' workload and left them feeling somewhat burnt-out at the end of the year.

School 2 Overview: Higher Performing Elementary School

School 2 is consistently one of the highest performing elementary schools in Aldine ISD. At the beginning of the year, the principal at School 2 was perceived by her teachers as a hard worker with realistic expectations. However, at the end of the year, some teachers reported a cliquey school culture. One teacher shared, *"It's very biased. If you're not in the clique, or if you've had a problem with someone that's in the clique or with an administrator or someone thinks you're causing a problem or anything"* (School 2, Teacher 1). One possible explanation for this culture shift could be that teachers at School 2 received their final ratings right before my interviews, and many were not happy with their overall scores. Indeed, the principal had very high expectations and gave very few Level 4 ratings.

On average, teachers at School 2 had seven years of experience, with 41% of faculty in their first five years of the profession, which is similar to the district average. Given the school's strong record of high performance, there were several veteran teachers who had been teaching at the school for many years. Interviewed teachers had mixed opinions of the system, which also changed during the course of the year. I interviewed three novices (two were new to the profession) and three veterans. One of the novice teachers was consistently positive (a sponge teacher), while the other two novices were overwhelmed by the system's expectations and burnt out by the end of the year. Of the

veterans, two started out the year invested in improving performance. However, by the end of the year, one was insulted by the system (given its perceived unfairness) and the other was more skeptical of the system's impact. The final veteran teacher was very skeptical at the beginning of the year, but over the course of the year, came to recognize that the Danielson Framework captured her impact (she was one of the few Highly Effective teachers at the school), which increased her appreciation for the system.

At the beginning of the year, there was some skepticism toward the system, though most teachers understood the system as a way to improve teacher performance. One teacher shared that INVEST was *"a good program to see what teachers were doing in the classroom, what their strengths were, what their weaknesses, because sometimes teachers think they are doing an awesome job, but it's always good for someone else to come in and critique you and let you know there's some changes or whatever."* In contrast, some veteran teachers viewed it as designed to dismiss bad teachers. According to one veteran teacher, *"At first, myself, I was like, like most people in the teaching profession, thinking it is a tool to get rid of teachers or to make it harder for them to achieve higher standards."*

At the end of the year, teachers' attitudes towards INVEST were mixed. Some teachers had a positive interpretation of INVEST as a system to improve student growth. One teacher shared that *"the whole purpose of it was not to penalize teachers, but to make the students better."* Conversely, many teachers seemed to have a more negative perception of the system than they did at the beginning of the year. Teachers reported that the expectations of INVEST were too high and unattainable. As noted above, one teacher

also shared that there was bias in the system due to a cliquey nature. Another teacher identified an observation bias depending on the timing of the observation: *"On Mondays, everything is introduction, so I guess if you get observed on a Monday, you're going to score lower than if you get observed on a Wednesday or Thursday. And that's what happened to me."* This perception of bias could be attributed to the fact that teachers had different observers who provided varying levels of feedback along with different ratings, and that many teachers did not reach their desired level of performance.

School 3 Overview: Lower Performing Intermediate School

School 3 is an intermediate school (with only grades 5-6) and was identified as underperforming in previous years but met standards during the pilot year. The school's cohesive leadership team created a friendly and positive atmosphere at the school. School 3 had a strong culture of collaboration and clear expectations. Teachers reported that the principal used staff development to discuss new initiatives, which helped build teachers' understanding and trust with the administration.

On average, teachers at School 3 had 10 years of experience, with 35% of the faculty in their first five years of the profession. The teachers at School 3 had a strong sense of professional community and many reported the value of their grade level or learning communities. The majority of teachers I interviewed at School 3 had positive perceptions of the new INVEST system. Two of the three novice teachers were sponges and remained very enthusiastic about implementing changes in their practice over the course of the year, while the final novice teacher was overwhelmed by the expectations of the new system and looking for other employment. All of the veteran teachers were

committed to improving on the new system, though two were a bit more skeptical about specific aspects (in particular, the increased workload).

Most teachers saw the purpose of INVEST as improving teachers' performance and giving them the tools they need to succeed. At the beginning of the year, one teacher shared: *"I guess the purpose of it was to kind of pinpoint the needs in the classroom as far as the student growth and teacher growth."* The principal viewed INVEST as a vast improvement on PDAS; in particular, he found the clarity of expectations and streamlined timeline of INVEST helpful, both for teachers and for himself. He shared: *"With INVEST, what I like about it is that well, it breaks it down for you. So it kind of takes some of the guessing out of it."* He used this structure to build strong systems to help him manage implementation of the system, and as a result, teachers had a clear picture of the purpose and structure of INVEST. As one teacher shared, *"I always knew the purpose was going to be to improve teacher performance."*

Teachers' perspective on INVEST remained relatively consistent over the course of the school year, and teachers still felt positive about the system at the end of the year. In fact, many teachers reported that they had a better understanding of INVEST due to the consistent staff development: *"Now I understand it way better because we've had so much training on it, but the attitude towards it, no, I still feel it's pretty good and it's very descriptive as to where to improve and where you are doing good, so I like it."* Indeed, the principal had strong systems in place, both to complete observations and conversations and support each teacher on the system. For example, he organized each teacher's INVEST documents in a binder along with a schedule for development

detailing when specific INVEST skills would be addressed. As a result, he was able to build teacher buy-in and understanding of the system and most teachers felt supported, not intimidated, by administrators' observations. However, teachers shared similar concerns over the increased paperwork load caused by INVEST.

School 4 Overview: Higher Performing Intermediate School

School 4 is the highest performing intermediate school and one of the highest performing overall schools in Aldine ISD. During the pilot year, there was an influx of students moving into the catchment area, which placed a strain on school operations. Despite serving an increased student population, School 4 had a positive school climate given the principal's strong reputation among his teachers as being a fair and approachable leader. However, the same did not hold true for the assistant principals, and teachers reported some concern over variation in observation scores across administrators. Teachers had a strong sense of professional community, and the skills specialist noted that the faculty was collaborating even more as a result of INVEST.

On average, teachers at School 4 had 13 years of experience, with 23% of their faculty in their first five years of the profession. Given the school's strong record of high performance, there were many veteran teachers who had been teaching at the school for several years, and there was a distinct difference between the veteran and novice teachers' perceptions. Two of the three novices were sponges, who were receptive to feedback and motivated to improve their performance on the new system. In contrast, the three veteran teachers were more skeptical about the new system and had yet to make changes in their practice. One veteran teacher was insulted by the new system because

she believed she deserved a higher rating, and the assistant principal reported that this perspective was fairly common among the more experienced teachers.

At the beginning of the year, teachers were unsure about the purpose of INVEST. Several perceived the system as top-down, coming from *"higher-up,"* indicating an initial lack of buy-in at the school level. The assistant principal noted that INVEST provided teachers with all the tools they needed to support their development, yet it had been challenging to find the time to have professional development. She said, *"Breaking it down, I think that's what they need…but finding the time to help them go through it is the problem."* She also shared that some teachers were overwhelmed by the unfamiliarity of the new system, which was being implemented the same year as the new state test (STAAR): *"They just don't know. And I think that's been the biggest thing with all of this is the unknown. They're used to knowing, ok, this is what I have to do…So between the STAAR and the INVEST, it's been challenging."*

By the end of the year, as a result of increased professional development, teachers had built their understanding of the system and had begun to see INVEST as a system designed to help both teachers and students improve performance. Teachers also felt they were given an opportunity to "prove themselves" through conversations with their administrators during the pre- and post-conferences, which helped them view the tool as a way to self-reflect and drive their own improvements in practice. However, several veteran teachers remained frustrated by not attaining the highest level of performance on the new system, which they were accustomed to receiving on the old system (PDAS). As

the assistant principal shared, *"we have winners in our building who are all leaders and they all want to be distinguished."*

School 5 Overview: Lower Performing Ninth Grade School

School 5 is traditionally an underperforming 9th grade school; however, it met standards during the pilot implementation year for the first time in several years. According to the principal and many of the teachers, School 5 serves a particularly high need and challenging student population. Teachers reported that the school can be a bit disorganized which often leads to difficulty implementing new systems with fidelity. However, the principal had a strong reputation for being a well-respected leader who communicates regularly with her staff.

On average, teachers at School 5 had 12 years of experience, with 38% of their faculty in the first five years of the profession. Of the six teachers I interviewed at School 5, three teachers were new to the school (with two new to teaching altogether) and three had been at School 5 for quite a few years. Of the novices, the two who were new to teaching were generally receptive to the new system but reported feeling overwhelmed and burnt out by the end of the year. The veterans' perspectives varied, though the majority was skeptical about specific aspects of the system. While this skepticism led several to feel frustrated by the system, others remained more apathetic. For the apathetic veterans, INVEST was merely another system that they did not find particularly relevant to their practice.

At the beginning of the year, INVEST was communicated as a way to improve teacher performance, but many teachers were skeptical that the system might simply be a

way to remove poor performing teachers. One teacher shared: *"My biggest concern is you'll have administrators that will want to use that as a kind of gotcha in the whole thing instead of them using it as what the district says it's supposed to be for, which is helping teachers improve."* Overall, teachers reported feeling nervous and overwhelmed by the new system.

Over the course of the year, the principal struggled to maintain the timeline and as a result, INVEST was not implemented with fidelity. New teachers reported appreciating the feedback they received and indicated that it helped them reflect on their practice and identify a pathway towards improvement. However, several veteran teachers had not received feedback, and none of the teachers had completed their summative conference. Though the timeline was difficult to maintain during the pilot year, the principal believed strongly in INVEST's ability to initiate positive changes in teaching practice. She shared the following: *"I believe in the Danielson Rubric, I think it's amazing. I think it will help build and grow great teachers."* Though the principal was confident she could help teachers achieve Level 4 performance, most teachers felt that Level 4 had been communicated as being unattainable, particularly at the high school level. As the year progressed, teachers adapted to new expectations of the system and were able to reflect on their own practice using the Danielson rubric. However, teachers reported working harder than in years past without being recognized for it, which lowered morale.

School 6 Overview: Higher Performing Ninth Grade School

School 6 has historically been one of the flagship schools in Aldine; however, in the year of the pilot, it did not meet state standards because it fell short on measures of

postsecondary readiness. The school has a very structured environment with a distinct culture, which has led to differing perceptions among the teacher corps. Half of the interviewed teachers reported that they really appreciated the structure and that the orderly environment made it easier to focus on instruction. The other half were incredibly frustrated and believed the leadership imposed unreasonable expectations and was generally unsupportive of teachers' efforts. In particular, several teachers complained that they were required to do paperwork that was not purposeful and kept them from meeting their students' needs.

On average, teachers at School 6 had eight years of experience, with 60% of the faculty in their first five years in the profession. The teachers I interviewed at School 6 had quite different perceptions of their experience, which subsequently affected how they viewed INVEST. Two of the six teachers (one novice, one veteran) were incredibly frustrated by the new system; indeed, one of them wrote a memo titled, *"Why INVEST is a train wreck of epic proportions that needs to be obliterated for the good of all mankind."* The other two novice teachers were receptive and felt the system provided a good structure for improving their instruction, while the remaining veteran teachers were neutral though somewhat skeptical of particular aspects of the new system. For purposes of this analysis, teachers' commentary will be categorized as either the "negative" or "neutral" group.

Among both the negative and neutral groups of teachers, INVEST was perceived as a new accountability system designed to evaluate teacher performance. One teacher shared that most teachers saw the purpose of INVEST as *"evaluating teachers and their*

performance in the classroom. Going to make sure that you are teaching the material and your content, you're managing your students, you're not taking a free pass every day." While the neutral group of teachers believed some level of accountability was necessary and even desirable, the negative group thought the accountability system was used to blame teachers. One particularly disgruntled teacher shared, *"this is just being used to basically give people who think we are slaves as opposed to professionals a leg to stand on when they tell us, well, you need to form relationships with these students and if they're not showing up it's your fault"* (School 6, Teacher 2).

The principal struggled with implementation fidelity and reported that the timeline was difficult with so many increased demands and the number of new teachers on his campus, *"like I said, the only thing is the timelines. Just making sure that's thought through. I'm all about providing feedback to the teachers and to the staff. But you only have 24 hours in the day."* However, he strongly believed that the new system resulted in constructive feedback and useful dialogue, *"allowing individuals to grow, allowing individuals to really craft what they're doing and do it well."* At the end of the year, several teachers had not yet received feedback on their formal observations and the majority of teachers reported limited conversations about their practice. Additionally, the administration struggled to implement the new online Teachscape system, which appeared to contribute to the delay in feedback. When they did receive it, teachers reported that they appreciated feedback on their performance, but given implementation challenges, this was not as frequent as they would have desired.

Section Two: Key School Organizational Factors

As these school profiles make clear, teachers' reactions to INVEST varied considerably across schools. In this section, I used multiple regression analyses to examine which school organizational factors were most influential in predicting teachers' motivation, effectiveness, and retention on the new system in the population of teachers and then revisited these school profiles to help explicate findings. As discussed in Chapter 2 (and presented in Table 2-3), I used the following school climate indicators listed below (each of these measures is on a scale of 1-5 with 1 being strongly disagree and 5 being strongly agree):

- ***Quality of administration*** – teachers' perceptions of the quality of administration. The survey captured information on teachers' opinions on administrators' supportive behavior, ability to enforce student conduct, and communication about vision.
- ***Positive support*** – teachers' perceptions of the level of support. The survey captured information on teachers' opinions on support from parents, the availability of materials, and the support for work with challenging student populations.
- ***Level of control***– teachers' perceptions of their level of control and influence over their work. The survey captured information on teachers' control over the selection of content/topics, teaching techniques, and disciplining students.

- ***Presence of a Professional Community*** – teachers' perceptions of the presence of professional community at the school. The survey captured information on the level of cooperative effort and shared mission at the school.

Not surprisingly given the important role of the administration in setting the school culture, perceptions of the quality of administrative leadership were correlated with the other measures of school climate – in particular, positive support ($r = .53$, $p < .001$), the level control ($r = .46$, $p < .001$), and the presence of professional community ($r = .54$, $p < .001$). Table 6-2 summarizes the correlations between the various factors.

Table 6-2

Correlations between School Characteristics

	Mean (SD)	1	2	3	4
1. Quality of administration	3.80 (0.88)	-	0.53***	0.46***	0.54***
2. Positive support	3.18 (0.81)		-	0.43***	0.45***
3. Level of control	3.61 (0.77)			-	0.34***
4. Professional community	3.48 (0.79)				-

*$p < .05$. **$p < .01$. ***$p < .001$

Teachers' perceptions of school climate were also associated with their motivation and performance on the new system. As demonstrated in Table 6-3, as was the case with individual characteristics, these school characteristics tended to be more correlated with system motivation (expectancy and value), reported changes in practice, and burnout than with actual system performance or turnover from the district. Of the

Table 6-3

Correlation between School Characteristics and Teacher Outcomes

Measure	System Expect-ancy	System Value	Danielson Framework	Student Growth Percen-tiles	Reported Changes in Practice	Turnover	Burnout
Individual Characteristics							
Teaching Grit	0.29***	0.28***	0.09**	0.02	0.29***	-0.10*	-0.53***
Overall Grit	0.11***	0.17***	0.08*	0.02	0.10**	-0.04	-0.30***
Conscientious-Ness	0.09**	0.19***	0.16***	-0.02	0.03	-0.00	-0.21***
Extraversion	0.05	0.07*	-0.01	0.03	0.06*	0.00	0.00
Agreeableness	0.09**	0.14***	-0.07*	-0.04	0.14***	-0.01	-0.13***
Emotional Stability	0.09**	0.05	-0.11***	0.14*	0.02	-0.02	-0.27***
Openness	0.13***	0.12***	0.00	-0.07	0.10**	0.02	-0.05
School Characteristics							
Quality of administration	0.30***	0.25***	0.12***	0.04	0.29***	-0.14***	-0.39***
Positive support	0.26***	0.18***	0.11***	0.02	0.26***	-0.10**	-0.38***
Level of control	0.23***	0.22***	0.12***	0.08*	0.26***	-0.09**	-0.35***
Professional community	0.18***	0.15***	-0.04	0.02	0.27***	-0.05	-0.27***
System Characteristics							
Impact Factor	0.47***	0.33***	0.01	-0.04	0.52***	-0.11***	-0.33***
Observation Factor	0.44***	0.37***	0.09**	0.01	0.45***	-0.06	-0.24***
Understanding Factor	0.40***	0.41***	0.08*	0.02	0.46***	-0.09**	-0.27***
Support Factor	0.45***	0.35***	0.01	-0.02	0.55***	-0.10**	-0.25***
Student Growth Factor	0.50***	0.34***	0.05	-0.01	0.44***	-0.03	-0.31***
Goal Factor	0.38***	0.36***	0.03	-0.02	0.54***	-0.09**	-0.25***

Note: For all measures, N=1097, except for Student Growth Percentiles measure, where N=651. *$p < .05$. **$p < .01$. ***$p < .001$

school characteristics, the quality of the administration was the most positively correlated with teachers' belief in their ability to improve practice on INVEST (system expectancy), $r = .30$, $p < .001$, the value they placed on the new system (system value), $r = .25$, $p <$

.001, and their reported changes in practice, $r = .29$, $p < .001$, as well as the most negatively correlated with their level of burnout, $r = -.39$, $p < .001$, and turnover from the district, $r = -.14$, $p < .05$. In other words, teachers who had better perceptions of the quality of their administrative leadership were more likely to be motivated on INVEST, to report implementing changes in their practice, and to avoid becoming burnt out by their work. Teachers' perceptions of the level of positive support and control/influence over their work were also positively correlated with teachers' motivation and performance and negatively correlated with their burnout and turnover.

School Characteristics and Teacher Motivation

This section presents the same set of analyses and tables used in the individual characteristics section, but this time with a focus on how the first outcome of interest – teacher motivation – was influenced by school characteristics. Again, motivation was captured from two distinct (but related) sets of beliefs: an individual's expectancy (or belief in their ability) and an individual's value (or the importance associated with a particular task).In this section, I discuss which school characteristics were most influential in predicting teacher motivation on the new system and compare the school profiles to further elucidate these results.

Table 6-4

Regression Analysis Predicting System (INVEST) Expectancy

Measure	Without School Effects			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.28(0.03)***	0.17(0.04)***	0.08(0.04)*	0.25(0.04)***	0.15(0.04)***	0.08(0.04)
Overall Grit	-0.02(0.04)	-0.03(0.04)	-0.02(0.04)	-0.03(0.04)	-0.04(0.04)	-0.02(0.03)
Conscientious -ness	-0.05(0.04)	-0.03(0.05)	0.00(0.04)	-0.05(0.04)	-0.03(0.04)	-0.01(0.04)
Extraversion	0.03(0.03)	0.04(0.03)	0.04(0.03)	0.01(0.03)	0.02(0.03)	0.03(0.03)
Agreeableness	0.01(0.04)	-0.01(0.04)	-0.01(0.03)	0.02(0.04)	-0.00(0.04)	-0.01(0.04)
Emotional Stability	0.00(0.03)	0.00(0.03)	-0.03(0.03)	-0.00(0.03)	-0.00(0.03)	0.04(0.03)
Openness	0.09(0.04)*	0.09(0.04)*	0.09(0.04)*	0.10(0.04)**	0.09(0.04)*	0.08(0.04)*
School Characteristics						
Quality of Administration		0.16(0.05)**	0.08(0.05)		0.14(0.05)**	0.08(0.05)
Positive Support		0.09(0.06)	0.03(0.05)		0.11(0.05)*	0.02(0.05)
Level Control		0.04(0.04)	-0.01(0.04)		0.06(0.04)	-0.00(0.04)
Professional Community		-0.03(0.04)	-0.07(0.04)		-0.02(0.04)	-0.05(0.04)
System Characteristics						
Impact Factor			0.05(0.09)			0.03(0.09)
Observation Factor			0.12(0.06)*			0.08(0.06)
Understanding Factor			-0.14(0.07)			-0.09(0.07)
Support Factor			0.28(0.09)**			0.25(0.09)**
Student Growth Factor			0.37(0.10)***			0.36(0.09)***
Goal Factor			-0.10(0.09)			-0.07(0.10)

Note: All continuous variables have been standardized. *$p < .05$. **$p < .01$. ***$p < .001$

School Characteristics and Teacher Motivation

Teachers' perceptions of their school climate influenced their motivation on INVEST. Since the principal was the vehicle for implementing INVEST, we would expect the quality of administrative leadership to influence teachers' motivation on the new system. Table 6-4 Model 2 showed that teachers' reports of administrative

Table 6-5

Regression Analysis Predicting System (INVEST) Value

Measure	Without School Effects			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.26(0.04)***	0.17(0.05)**	0.09(0.04)*	0.24(0.04)***	0.15(0.04)**	0.09(0.04)*
Overall Grit	0.03(0.04)	0.03(0.04)	0.03(0.04)	0.03(0.04)	0.03(0.05)	0.03(0.04)
Conscientious-ness	0.08(0.04)*	0.09(0.04)*	0.09(0.04)*	0.08(0.04)	0.08(0.04)*	0.08(0.04)*
Extraversion	0.05(0.03)	0.06(0.02)	0.04(0.03)	0.04(0.03)	0.05(0.03)	0.05(0.03)
Agreeableness	0.06(0.04)	0.04(0.04)	0.04(0.04)	0.05(0.04)	0.03(0.04)	0.04(0.04)
Emotional Stability	0.04(0.04)	-0.04(0.03)	-0.03(0.03)	-0.02(0.04)	-0.01(0.04)	-0.01(0.03)
Openness	0.02(0.04)	-0.00(0.04)	0.01(0.04)	0.04(0.04)	0.01(0.04)	0.02(0.04)
School Characteristics						
Quality of Administration		0.12(0.05)*	0.06(0.04)		0.15(0.05)**	0.08(0.05)
Positive Support		0.04(0.04)	-0.02(0.04)		0.05(0.05)	-0.05(0.04)
Level of Control		0.05(0.03)	0.02(0.03)		0.07(0.03)	0.02(0.03)
Professional Community		0.01(0.04)	-0.01(0.04)		-0.01(0.05)	-0.04(0.05)
System Characteristics						
Impact Factor			-0.04(0.09)			-0.00(0.09)
Observation Factor			0.12(0.06)*			0.13(0.05)*
Understanding Factor			0.21(0.07)**			0.22(0.08)**
Support Factor			0.07(0.08)			0.08(0.09)
Student Growth Factor			0.01(0.10)			0.02(0.11)
Goal Factor			0.10(0.07)			0.09(0.07)

Note: All continuous variables have been standardized. *$p < .05$. **$p < .01$. ***$p < .001$.

leadership significantly influenced their level of system expectancy (or belief in their ability to perform well on the system), $B = .16$, $t(908) = 10.24$, $p < .01$, and Model 5 (in Table 6-4) revealed that this was robust to the inclusion of school effects. In addition to positively influencing system expectancy, Table 6-5 Model 2 confirmed that teachers' perception of their leader was also a significant predictor of the value they placed on

performing well on the system, $B = .12$, $t(908) = 9.89$, $p < .05$ and again, Model 5 validated that this effect was still significant with school effects included in the model. As demonstrated in Table 6-3, other school organizational factors (besides principal leadership) were also associated with teachers' system motivation. However, when included in the regression analysis, the level of support was the only other factor predicting either system expectancy or value. This suggests that the quality of administration at least partially explains the relationship between teachers' perceptions of other working conditions and their motivation.

This section will explore which aspects of principal leadership mattered most for teachers' motivation by comparing two schools – School 1 and School 2. In School 1, the principal used the Danielson Framework to empower teachers to drive their own improvement process, and overall, teachers stayed motivated to improve performance on INVEST. In School 2, the principal rigorously implemented the system, yet teachers had mixed attitudes toward INVEST, with many expressing concern over the unattainability of Level 4 performance. Interview data suggested that differences in motivational responses resulted from several key factors.

Principal communication and vision-setting. To be motivated to improve practice, teachers need to believe change is essential and understand how INVEST will support their growth. In School 1, the principal set aspirational yet realistic expectations for system implementation at the beginning of the year. She consistently communicated to teachers that the ultimate goal of the new system was to support teachers' performance so they could impact students, which tapped into their intrinsic value as educators. However,

she was also realistic with her expectations and communicated to her teachers that this was a "learning process" and that they were going to "grow together." Since she was overwhelmed by the timeline herself, she was empathetic to teachers' concerns about workload and was careful to scaffold training and provide time for collaborative planning during the day. In School 2, the principal also believed INVEST was designed to support teachers, but her initial messaging to teachers did not reflect this understanding. She warned teachers that the new system was coming in a joking manner:

> I went and told them from the very beginning, OK guys, I'm just letting you know, something's coming from the district. It's called INVEST. I would joke around and I would say, you know, I keep on hiding, but I'm going to let you know, there aren't that many principals. There is only a team of principals that's part of this whole thing, so more than likely I really think they're going to be using those principals to go ahead and pilot it. So I told them from the very beginning when I got involved in it. And once we got picked, I went ahead and said, OK, I told you all. They caught me. I couldn't hide any longer. It was kind of a big joke.

Consequently, though teachers at School 2 were initially fairly open to INVEST, they viewed it less as a tool for improvement and more as a way to "measure how well you are teaching " (Teacher 4), "appraise you and see how you are teaching" (Teacher 6), and "see what administrators are looking for" (Teacher 2). This meant that teachers' system value did not necessarily connect with their own intrinsic value as educators (as it had in School 1).

Principal empowerment of teachers. Teachers were more likely to be motivated by the new system when it provided them with ownership over their practice. In School 1, the principal used the rubric to guide teachers through a self-directed learning process by asking them to reflect on their practice, *"I would ask them, why do you think I gave*

you a two in this area and they were able to tell me…. Where they are proficient I always ask them what do you think you could have done that would move you to distinguished." Teachers reported that the pre-conference was particularly important because it gave them a chance to share what they were doing in the classroom and guide the observation. Given their involvement, teachers reported that the process felt more like self-reflection than evaluation. In School 2, the system was implemented with fidelity and teachers reported having conversations which were very specific and focused on areas for improvement. One teacher shared how conversations were typically structured: *"She told me exactly what I needed to have. She used the rubric. It was right in front of her. It was straightforward.*" Though teachers appreciated the feedback, conversations were more administrator-driven, which left teachers feeling like they did not have as much control over the process. Compared with School 1, this contributed to lower levels of value.

School Characteristics and Teacher Effectiveness

Once teachers are initially motivated to improve performance, their ability to sustain improvements in practice is, in part, a function of their working environment. According to the theory of deliberate practice, teachers need targeted, immediate, and consistent feedback to enhance performance. This section examines how school characteristics influenced teachers' effectiveness on the new system, as measured by the Danielson Framework for Teaching (the observation measure). I present the same set of analyses discussed in the previous section on individual characteristics, with a new focus on school characteristics.

Table 6-6

Regression Analysis Predicting Teacher Effectiveness on Danielson Observation Measure

Measure	Without School Effects			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.09(0.04)*	0.03(0.05)	0.03(0.04)	0.12(0.04)**	0.07(0.05)	0.06(0.04)
Overall Grit	0.01(0.05)	0.00(0.04)	0.01(0.04)	0.00(0.04)	-0.01(0.04)	-0.00(0.04)
Conscientious-ness	0.15(0.04)***	0.15(0.04)**	0.14(0.04)***	0.15(0.04)***	0.14(0.04)***	0.13(0.04)**
Extraversion	-0.02(0.04)	-0.02(0.03)	-0.02(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)
Agreeableness	-0.10(0.03)**	-0.10(0.03)**	-0.10(0.03)**	-0.10(0.03)**	-0.09(0.03)**	-0.09(0.03)**
Emotional Stability	-0.12(0.03)**	0.12(0.03)***	0.11(0.03)***	-0.09(0.03)**	-0.09(0.03)**	-0.09(0.03)**
Openness	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.02(0.03)
School Characteristics						
Quality of Administration		0.11(0.04)*	0.08(0.05)		0.12(0.04)**	0.09(0.04)*
Positive Support		0.07(0.03)*	0.08(0.04)*		0.07(0.03)	0.07(0.04)
Level of Control		0.09(0.03)**	0.07(0.03)*		0.10(0.03)**	0.08(0.03)*
Professional Community		-0.14(0.04)***	-0.12(0.04)**		-0.19(0.04)***	-0.16(0.05)**
System Characteristics						
Impact Factor			-0.07(0.08)			-0.07(0.08)
Observation Factor			0.11(0.06)*			0.09(0.06)*
Understanding Factor			0.13(0.06)*			0.18(0.07)*
Support Factor			-0.16(0.10)*			-0.15(0.09)*
Student Growth Factor			-0.03(0.08)			-0.02(0.08)
Goal Factor			0.03(0.06)			-0.01(0.07)

Note: All continuous variables have been standardized. *$p < .05$. **$p < .01$. ***$p < .001$

As was the case with teacher motivation, the quality of administration again emerged as the most important influence on teachers' effectiveness on the new system. As demonstrated in Table 6-6, Model 2, quality of administration, support, and control over practice all emerged as positive and significant predictors of teachers' scores on the Danielson Framework. In Model 2, administration had the largest positive coefficient (*B*

= .11, $t(886) = 8.14$, $p < .05$) followed by control ($B = .09$, $p < .01$) and support ($B = .08$, $p < .05$) and with the exception of level of support, these coefficients were robust to the inclusion of school effects in Model 5. Interestingly, the level of professional community was a negative predictor of Danielson scores, $B = -.14$, $t(886) = 8.14$, $p < .001$, which was also significant with the inclusion of school effects.

To explicate these results, I explored how features of principal leadership and school climate influenced teachers' system improvement by comparing two schools – School 3 and School 5. In School 3, the principal scaffolded the introduction of the new system and provided consistent support, which led teachers to feel supported on the new system. In School 5, INVEST was not implemented with fidelity, which meant that teachers did not receive the necessary feedback to improve performance. Interview data suggested that these differences in teachers' ability to increase effectiveness resulted from several school-level factors discussed below.

Principal conscientiousness. In line with the theory of deliberate practice, quality feedback should diagnose specific needs and offer immediate and explicit strategies for improving instruction. Principals' ability to manage the complexity of the new INVEST system was influenced by their attention to detail, as well as their organizational and time management skills. At School 3, all teachers reported that they received specific and detailed feedback from their appraisers: *"It had comments on just about every domain that I got observed on and he had a level on it. He gave me a breakdown of everything."* (School 3, Teacher 4). Further, this feedback was "simple and quick" focusing on discrete pieces of the lesson and what teachers could do to make immediate changes in practice.

Conversely, at the end of the year, most teachers at School 5 had yet to receive feedback from their observations and the feedback that was received was fairly general, focusing on instituting "group-work" or "conducting student-driven lessons." Unlike the principal at School 3, the principal at School 5 reported being overwhelmed by the new system and struggled to conduct observations and provide feedback in a timely manner.

Systems for support and reflection. Performance can only be improved when there are structures in place to support ongoing reflection and improvements in practice. At School 3, weekly staff meetings were focused on different components of the Danielson Framework (aligned to observation results) and skills were scaffolded over the course of the year based on difficulty and teacher need. Further, the principal intentionally structured conversations to gradually develop teachers' self-reflective capabilities. School 5's principal agreed that the developmental aspect of the new system was critical to the process. However, she did not have the same structures in place to ensure consistent implementation and reported being very overwhelmed by the expectations of the new system:

> We need to reduce that number of walkthroughs. Also, it's the Teachscape. For every walkthrough it takes an hour, literally, it takes an hour to put the stuff in Teachscape. OK, you put it in, you send it to the teacher, you got to wait for the teacher to accept it, review it. The time. We don't have the time.

As a result, teachers at School 5 reported turning to one another for support. Engaging in deliberate practice is incredibly challenging and at least initially, cannot be done alone. Indeed, principals played an integral role in supporting teachers through the process. Since the principal was such a critical component of implementation, this may help

explain why the presence of professional community did not have a positive influence on teachers' performance.

School Characteristics and Teacher Retention

To increase effectiveness, teachers must sustain improvements in practice over time and avoid becoming burnt out by their work. As the school profiles demonstrated, teachers on some campuses seemed more concerned about the increased workload and its potential impact on teacher retention. To evaluate which school-level characteristics

Table 6-7

Binary Logistic Regression Table Predicting Turnover

Measure	Without School Effects		With School Effects			
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.79(0.05)**	0.88(0.06)	0.83(0.10)	0.79(0.06)**	0.85(0.07)*	0.82(0.11)
Overall Grit	0.92(0.08)	0.94(0.08)	0.94(0.11)	0.92(0.08)	0.93(0.08)	0.91(0.13)
Conscientious-ness	1.18(0.16)	1.17(0.10)	1.19(0.17)	1.18(0.12)	1.17(0.11)	1.25(0.19)
Extraversion	0.92(0.06)	0.92(0.05)	0.99(0.11)	0.90(0.05)	0.90(0.05)	0.99(0.10)
Agreeableness	1.01(0.08)	1.04(0.08)	1.11(0.12)	1.01(0.08)	1.03(0.08)	1.10(0.13)
Emotional Stability	1.04(0.09)	1.05(0.07)	0.96(0.12)	1.09(0.08)	1.10(0.12)	1.01(0.13)
Openness	1.04(0.09)	1.04(0.09)	1.09(0.15)	1.05(0.09)	1.05(0.09)	1.11(0.16)
School Characteristics						
Quality of Administration		0.82(0.06)**	0.68(0.07)**		0.81(0.06)**	0.70(0.08)**
Positive Support		0.98(0.08)	0.97(0.10)		0.94(0.08)	0.86(0.09)
Level of Control		0.89(0.06)	0.91(0.12)		0.90(0.06)	1.00(0.15)
Professional Community		1.04(0.07)	1.11(0.12)		1.15(0.08)	1.22(0.32)
System Characteristics						
Impact Factor			0.41(0.15)**			0.40(0.14)*
Observation Factor			0.98(0.15)			1.02(0.17)
Understanding Factor			0.85 (0.19)			0.87(0.20)
Support Factor			1.32(0.32)			1.22(0.32)
Student Growth Factor			3.13(1.07)**			3.38(1.30)**
Goal Factor			0.92(0.26)			0.93(0.27)

Note: All continuous variables have been standardized. *$p < .05$. **$p < .01$. ***$p < .001$

influenced teachers' turnover, I present the same analysis discussed in the individual characteristics section, with a focus on school characteristics, below.

Teachers who chose to stay at their school had more favorable perceptions of the school's administration and working conditions across all survey measures. As demonstrated in Table 6-7, Model 2, the quality of administration was the only school-level variable to emerge as a significant predictor of teachers' retention. Even when including school effects and system characteristics in the model, Model 6 illustrates that teachers who were one standard deviation higher in perceptions of administrator quality were 30% less likely to leave teaching in the district, $OR = .70$, $p < .01$.

This section will explore which features of principal leadership appeared to influence teachers' level of burnout and retention decisions at two schools – School 4 and School 6. At School 4, while some teachers appreciated the principals' strict requirements, it contributed to higher rates of burnout among a subset of teachers on the campus. At School 6, the principal built strong trusting relationships with his teachers. Rather than use INVEST to impose expectations, he empowered teachers to use the rubric to drive their own reflective process. This level of trust and recognition helped teachers avoid burnout and remain committed to teaching.

Trust. Teachers' level of trust in their principal influenced their thinking about whether to stay teaching in Aldine. School 4 teachers had a great deal of respect for the principal and found him to be both caring and encouraging of their development. As one teacher put it, "*They all like him and he likes all of them. He's very laid back. He's very calm, has an aura*" (School 4, Teacher 2). Teachers felt that Principal 4 was focused on

supporting their practice and always available to discuss challenges in a non-threatening manner, which meant INVEST wasn't viewed as a "gotcha" system. Though he already took this approach prior to the pilot, Principal 4 shared that INVEST had further supported his work with teachers by helping him "to bond with the teachers, to talk to them, to get them to share."

In contrast, teachers' perceptions of their administration at School 6 ranged significantly. Some teachers appreciated the principal's no nonsense leadership style, which led to a disciplined school environment. One such teacher shared, "*Here at [School 6], everything is extremely structured, disciplined, and the students are afraid of Principal 6 so that makes a big difference because you don't want to go to the office"* (School 6, Teacher 5). In contrast, other teachers found the principal to be unsupportive and felt pressured by the rigid school culture. One of these teachers explained, *"I don't even bother saying anything anymore because I know that my idea will be shot down"* (School 6, Teacher 2). Additionally, this teacher perceived the principal at School 6 as being dismissive, "*He asked me how I was doing and I said I had been under much pressure and stress and he basically said I need to learn to manage my time better."* Teachers reported that Principal 6's polarizing leadership style contributed to the school's lower retention rate and a revolving door of new teachers every year.

Recognition. Teachers appeared more likely to avoid burnout when the new system aligned with their personal values and helped them to recognize their impact on students. At School 4, the principal helped teachers feel recognized for their

accomplishments by hosting pre-conferences in their classroom or "in their element" as he phrased it:

> One of the things that I think is neat for us is I'm doing the pre-conferences in their room. Instead of them coming to me and bringing all their stuff, I decided, you know what, to relieve a little bit of stress, we're scheduling these things, the administrators go and schedule a time they can go to the classroom. So we can be there, we're in the teachers' element, and while we're discussing things, we're in the classroom.

At these conferences, teachers were able to showcase their work and ensure that the new system aligned with their own personal goals. As one teacher noted, "*it allowed me to know where I need to work personally, my personal goals as a teacher and how successful I'm going to be delivering the instruction to students*" (School 4, Teacher 3). Teachers at School 4 reported that they were more likely to stay in teaching because their principal really took the time to understand – and recognize– their work and INVEST helped them see their impact, which made the additional work worth the effort. In contrast, School 6 teachers reported that INVEST felt "dictated from the top" and "administrator-driven." Instead of sharing artifacts in their classroom, they were required to assemble an "access to excellence" binder with a set of mandated resources so the principal could monitor their activities and ensure they were meeting expectations. Rather than feeling recognized for their work, School 6 teachers reported that the process was nothing more than "unnecessary paperwork" or "redundant crap." Though the process alone was not enough to drive them out of teaching, they believed INVEST compounded an already punitive culture that made teaching less rewarding.

Summary

In sum, even when taking individual characteristics into account, teachers' responses varied considerably across schools. We already know from Chapter 3 that teachers in high schools and on higher performing campuses appeared to have lower perceptions of INVEST. This chapter extended this analysis by exploring differences in teachers' attitudes across the six case study schools. School 1 (underperforming elementary) had a collaborative culture and teachers placed strong trust in the principal because she gave them considerable control over their practice. In School 2 (higher performing elementary), the principal had very high expectations for performance, and some teachers expressed concerns about a cliquey nature and unreasonable standards. School 3's (underperforming intermediate) principal used staff development time to discuss the new system which helped build teachers' understanding and trust in INVEST. School 4 (higher performing intermediate) had a positive climate which in large part was due to the principal's strong reputation among his teachers. School 5 (underperforming ninth grade) had a reputation for being disorganized and having a more skeptical community of veteran teachers. School 6 had a very structured environment with a distinct culture. Some teachers appreciated the orderly environment, while others were frustrated by what they perceived to be unrealistic expectations imposed by the administration as part of INVEST.

Of all the school characteristics, the quality of administrative leadership was consistently the most important influence on teachers' outcomes. More effective principals (like the ones in Schools 1, 3, and 4) communicated a clear and compelling

vision for INVEST which was focused on professional growth and empowered teachers to drive their own improvement efforts. Additionally, these principals were not as overwhelmed by the new requirements and as such, effectively managed the implementation process to ensure that teachers received targeted, meaningful, and consistent feedback over the course of the year. As a result, they developed trusting relationships with their staff, which helped teachers feel recognized for their hard work. In contrast, less effective principals (like the ones in Schools 5 and 6) struggled to efficiently manage the new evaluation process, which meant teachers received limited feedback on their performance. Additionally, these principals failed to communicate that INVEST was about growth, and instead, teachers were more likely to view INVEST as a "gotcha" system designed to hold them accountable to unreasonable standards. Teachers' beliefs about the purpose and usefulness of INVEST certainly appeared to influence their motivation and performance. This variation in system perceptions will be more systematically explored in the subsequent chapter.

CHAPTER 7: SYSTEM-LEVEL VARIATION

To influence behavioral change, individuals must be motivated by the system itself. As Chapter 4 demonstrated, teachers' personal motivation differed from their motivation to perform well on INVEST. In other words, teachers could believe in their own abilities but question whether or not those abilities would be enough to meet system expectations and/or whether the system expectations were worth meeting in the first place. In addition to varying across individual and school level characteristics discussed in the prior two chapters, attitudes were also influenced by teachers' perceptions of the system itself.

As discussed in Chapter 2, I collected information on teachers' overall perceptions of evaluation (in both pilot and non-pilot schools) and then asked more specific questions on pilot teachers' perceptions of INVEST. Given this chapter's focus on INVEST, I analyzed the INVEST specific variables as a source of system variation. I conducted exploratory factor analysis so that I could explain the larger number of survey questions with a smaller set of latent constructs (discussed in Section One). After conducting factor analysis, I used these factors as variables in subsequent analyses to investigate how system-level characteristics predicted teachers' outcomes in the district when controlling for individual and school characteristics (discussed in Section Two).

Section One: System-Level Exploratory Factor Analysis

I employed exploratory factor analysis to determine which theoretical constructs underlay the 19 survey questions collected on INVEST (outlined in Table 2-3) and then examined the extent to which these constructs represented the original scales I had

developed for the analysis. To begin, I first used Varimax (orthogonal rotation) followed by Promax (oblique rotation) to better approximate simple structure. As demonstrated in Table 7-1, the analysis of INVEST attitudes yielded six factors, using the Kaiser criterion (Eigenvalue >1) and the Scree test. The factor loadings for each item are reported, as well as the uniqueness of each item and the Eigenvalue and total variance explained by each factor. These latent factors aligned closely to the scales I developed.

- ***Factor 1 (Impact Factor)*** represented teachers' perceptions of the possible positive impact of the new system due to the high loadings (> .4) by the following items – the system's impact on teaching, teacher development, and student growth, and overall. Factor 1 explained 69% of the total variance in the dataset.
- ***Factor 2 (Observation Factor)*** represented perceptions of the accuracy and fairness of the Danielson observation measure and process. Factor 2 explained 67% of the total variance in the dataset.
- ***Factor 3 (Understanding Factor)*** represented perceptions of the quality of communication and training and how this built system understanding. Factor 1 explained 60% of the total variance in the dataset.
- ***Factor 4 (Support Factor)*** represented perceptions of the quality of feedback and opportunities for professional growth and support under the new system. Factor 1 explained 52% of the total variance in the dataset.
- ***Factor 5 (Student Growth Factor)*** represented perceptions of the accuracy and fairness of the Student Growth Percentile Measure. Factor 1 explained 47% of the total variance in the dataset.

- ***Finally, Factor 6 (Goal Factor)*** captured information on the quality of the goal-setting process teachers went through as part of INVEST. Factor 1 explained 46% of the total variance in the dataset.

The uniqueness of the variables was small to moderate (.15 to .48) with one fairly unique variable (thoughts on the online Teachscape system) which did not load onto any of the factors.

Table 7-1

Exploratory Factor Analysis: INVEST-Specific Attitudes

Variable	Loadings						
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Unique-ness
Understanding							
Initial understanding	0.09	0.02	**0.66**	0.06	-0.07	-0.00	0.46
Ongoing communication	0.03	0.07	**0.72**	0.01	0.04	-0.02	0.42
Observation training	0.09	0.06	**0.67**	-0.04	-0.02	0.05	0.37
Teachscape online system	0.05	0.04	0.27	0.11	0.02	-0.08	0.75
Student growth training	-0.05	0.01	**0.57**	-0.06	0.15	0.07	0.48
Goal-setting							
Goal-setting focused efforts	0.10	0.10	0.27	-0.01	-0.08	**0.48**	0.41
Set challenging goals	0.19	0.04	0.02	0.07	0.05	**0.52**	0.42
Accuracy of INVEST Measures							
Overall Danielson	0.14	**0.53**	0.09	-0.00	0.18	0.01	0.28
Danielson Domain 1	0.05	**0.87**	0.04	0.01	-0.03	0.01	0.18
Danielson Domain 2	0.10	**0.91**	-0.03	-0.03	-0.10	0.02	0.17
Danielson Domain 3	0.03	**0.84**	0.02	0.02	-0.00	-0.01	0.18
Danielson Domain 4	0.02	**0.81**	0.02	0.02	0.08	0.00	0.25
Student growth percentiles	0.24	-0.02	0.04	0.02	**0.41**	-0.01	0.62
Growth & Improvement							
Quality of feedback	0.23	0.13	0.08	**0.48**	-0.03	0.02	0.34
Level of support	0.36	0.04	-0.04	**0.46**	0.05	0.09	0.27
Impact on my teaching	**0.86**	0.04	0.01	0.02	-0.05	0.07	0.16
Impact on teacher development	**0.79**	0.11	0.11	0.05	-0.00	-0.08	0.18
Impact on student growth	**0.85**	0.02	0.00	-0.03	0.07	0.03	0.15
Overall impact	**0.74**	0.11	-0.00	-0.03	0.10	-0.00	0.18
Eigenvalue	8.23	7.99	7.18	6.20	5.54	5.45	
% of variance	69.17%	67.12%	60.31%	52.11%	46.55%	45.76%	

Section Two: Key System Characteristics

Given the relationship between INVEST's features discussed in Chapter 3 (for example, teachers' understanding appeared to influence their perceptions of the purpose and impact of the system), we would expect these factors to be correlated. As demonstrated in Table 7-2, these correlations were, in fact, relatively large in magnitude. Though teachers had varying opinions about specific features of INVEST, this analysis suggests that their perceptions were significantly associated with each other.

Table 7-2

Correlations between System Characteristics

Measures	1	2	3	4	5	6
1. Impact Factor	-	0.68***	0.71***	0.81***	0.87***	0.76***
2. Observation Factor		-	0.72***	0.74***	0.74***	0.69***
3. Understanding Factor			-	0.78***	0.74***	0.78***
4. Support Factor				-	0.72***	0.84***
5. Student Growth Factor					-	0.66***
6. Goal Factor						-

**p < .05. **p < .01. ***p < .001*

In addition, teachers' perceptions of system features were associated with their motivation, performance, and turnover. As demonstrated in Table 7-3, system perceptions were more correlated with teachers' system expectancy (or belief in their ability to meet standards on INVEST) than their system value (or importance they placed on INVEST). However, the correlations with system characteristics were greater for both sources of motivation than they were with either individual or school characteristics. Additionally, teachers' perceptions of system characteristics were highly correlated with reported

changes in practice. One of the limitations of the Danielson measure is that we do not know how teachers performed at the beginning of the year, so it is not possible to track progress over the course of the year. Based on the large correlations between specific system features and reported changes in practice, we might expect system features to be more associated with improvements on the Danielson rubric than with actual scores. Table 7-3 also demonstrated that teachers who chose to stay in the district had more favorable perceptions of INVEST, particularly when it came to the perceived impact and support provided by the new system. Additionally, quality positive support, level of

Table 7-3

Correlation between System Characteristics and Teacher Outcomes

Measure	System Expectancy	System Value	Danielson Framework	Student Growth Percentiles	Reported Changes in Practice	Turnover	Burnout
Individual Characteristics							
Teaching Grit	0.29***	0.28***	0.09**	0.02	0.29***	-0.10*	-0.53***
Overall Grit	0.11***	0.17***	0.08*	0.02	0.10**	-0.04	-0.30***
Conscientiousness	0.09**	0.19***	0.16***	-0.02	0.03	-0.00	-0.21***
Extraversion	0.05	0.07*	-0.01	0.03	0.06*	0.00	0.00
Agreeableness	0.09**	0.14***	-0.07*	-0.04	0.14***	-0.01	-0.13***
Emotional Stability	0.09**	0.05	-0.11***	0.14*	0.02	-0.02	-0.27***
Openness	0.13***	0.12***	0.00	-0.07	0.10**	0.02	-0.05
School Characteristics							
Quality of administration	0.30***	0.25***	0.12***	0.04	0.29***	-0.14***	-0.39***
Positive support	0.26***	0.18***	0.11***	0.02	0.26***	-0.10**	-0.38***
Level of control	0.23***	0.22***	0.12***	0.08*	0.26***	-0.09**	-0.35***
Professional community	0.18***	0.15***	-0.04	0.02	0.27***	-0.05	-0.27***
System Characteristics							
Impact Factor	0.47***	0.33***	0.01	-0.04	0.52***	-0.11***	-0.33***
Observation Factor	0.44***	0.37***	0.09**	0.01	0.45***	-0.06	-0.24***
Understanding Factor	0.40***	0.41***	0.08*	0.02	0.46***	-0.09**	-0.27***
Support Factor	0.45***	0.35***	0.01	-0.02	0.55***	-0.10**	-0.25***
Student Growth Factor	0.50***	0.34***	0.05	-0.01	0.44***	-0.03	-0.31***
Goal Factor	0.38***	0.36***	0.03	-0.02	0.54***	-0.09**	-0.25***

Note: For all measures, *N*=1097, except for Student Growth Percentiles measure, where *N*=651. *$p < .05$. **$p < .01$. ***$p < .001$

understanding, and teachers' engagement in goal-setting were also negatively correlated with burnout and turnover, though to a slightly lesser extent.

System Characteristics and Teacher Motivation

As discussed in Chapter 3, teachers' attitudes toward INVEST were influenced by system design features and the implementation process. When teachers believed INVEST was designed to support their growth, they were more invested in making changes in

Table 7-4

Regression Analysis Predicting System (INVEST) Expectancy

Measure	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.28(0.03)***	0.17(0.04)***	0.08(0.04)*	0.25(0.04)***	0.15(0.04)***	0.08(0.04)
Overall Grit	-0.02(0.04)	-0.03(0.04)	-0.02(0.04)	-0.03(0.04)	-0.04(0.04)	-0.02(0.03)
Conscientiousness	-0.05(0.04)	-0.03(0.05)	0.00(0.04)	-0.05(0.04)	-0.03(0.04)	-0.01(0.04)
Extraversion	0.03(0.03)	0.04(0.03)	0.04(0.03)	0.01(0.03)	0.02(0.03)	0.03(0.03)
Agreeableness	0.01(0.04)	-0.01(0.04)	-0.01(0.03)	0.02(0.04)	-0.00(0.04)	-0.01(0.04)
Emotional Stability	0.00(0.03)	0.00(0.03)	-0.03(0.03)	-0.00(0.03)	-0.00(0.03)	0.04(0.03)
Openness	0.09(0.04)*	0.09(0.04)*	0.09(0.04)*	0.10(0.04)**	0.09(0.04)*	0.08(0.04)*
School Characteristics						
Quality of Administration		0.16(0.05)**	0.08(0.05)		0.14(0.05)**	0.08(0.05)
Positive Support		0.09(0.06)	0.03(0.05)		0.11(0.05)*	0.02(0.05)
Level of Control		0.04(0.04)	-0.01(0.04)		0.06(0.04)	-0.00(0.04)
Professional Community		-0.03(0.04)	-0.07(0.04)		-0.02(0.04)	-0.05(0.04)
System Characteristics						
Impact Factor			0.05(0.09)			0.03(0.09)
Observation Factor			0.12(0.06)*			0.08(0.06)
Understanding Factor			-0.14(0.07)			-0.09(0.07)
Support Factor			0.28(0.09)**			0.25(0.09)**
Student Growth Factor			0.37(0.10)***			0.36(0.09)***
Goal Factor			-0.10(0.09)			-0.07(0.10)

Note: All continuous variables have been standardized. *Controlling for all individual-level demographic characteristics* . *$p < .05$. **$p < .01$. ***$p < .001$

practice. Additionally, when they viewed the evaluation measures as providing an accurate picture of their performance, they seemed more likely to believe in their abilities to increase effectiveness. Finally, when they received quality feedback, teachers had better overall perceptions of INVEST's potential to bring about meaningful change.

Table 7-5

Regression Analysis Predicting System (INVEST) Value

Measure	Without School Effects			With School Effects		
	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	**Model 6: *Individual, School, and System Variation***
Individual Characteristics						
Teaching Grit	0.26(0.04)***	0.17(0.05)**	0.09(0.04)*	0.24(0.04)***	0.15(0.04)**	0.09(0.04)*
Overall Grit	0.03(0.04)	0.03(0.04)	0.03(0.04)	0.03(0.04)	0.03(0.05)	0.03(0.04)
Conscientiousness	0.08(0.04)*	0.09(0.04)*	0.09(0.04)*	0.08(0.04)	0.08(0.04)*	0.08(0.04)*
Extraversion	0.05(0.03)	0.06(0.02)	0.04(0.03)	0.04(0.03)	0.05(0.03)	0.05(0.03)
Agreeableness	0.06(0.04)	0.04(0.04)	0.04(0.04)	0.05(0.04)	0.03(0.04)	0.04(0.04)
Emotional Stability	0.04(0.04)	-0.04(0.03)	-0.03(0.03)	-0.02(0.04)	-0.01(0.04)	-0.01(0.03)
Openness	0.02(0.04)	-0.00(0.04)	0.01(0.04)	0.04(0.04)	0.01(0.04)	0.02(0.04)
School Characteristics						
Quality of Administration		0.12(0.05)*	0.06(0.04)		0.15(0.05)**	0.08(0.05)
Positive Support		0.04(0.04)	-0.02(0.04)		0.05(0.05)	-0.05(0.04)
Level of Control		0.05(0.03)	0.02(0.03)		0.07(0.03)	0.02(0.03)
Professional Community		0.01(0.04)	-0.01(0.04)		-0.01(0.05)	-0.04(0.05)
System Characteristics						
Impact Factor			-0.04(0.09)			-0.00(0.09)
Observation Factor			0.12(0.06)*			0.13(0.05)*
Understanding Factor			0.21(0.07)**			0.22(0.08)**
Support Factor			0.07(0.08)			0.08(0.09)
Student Growth Factor			0.01(0.10)			0.02(0.11)
Goal Factor			0.10(0.07)			0.09(0.07)

Note: All continuous variables have been standardized. *p < .05. **p < .01. ***p < .001.

These findings suggest that specific system features influenced teachers' system expectancy (or belief in their ability on the system) and system value (or the importance associated with the system). This section explores which system characteristics were most influential in predicting teacher motivation on the new system using the same

analyses discussed in Chapter 5 (individual characteristics) and Chapter 6 (school characteristics).

Understanding. During the pilot year of INVEST, teachers' value was influenced by their understanding of the new system and whether they felt it was intended to support their development. Table 7-4 Model 3 showed that the Understanding Factor (capturing teachers' reported understanding of the system and quality of communication and training) was a significant predictor of system value, $B = .21$, $t(908) = 12.23$, $p < .05$, and Model 6 (in Table 7-4) validated that this effect was still significant with school effects included in the model. Teachers who saw INVEST as designed for professional growth reported being more optimistic about their ability to control outcomes on the new system and generally, more positive about their abilities to improve practice. To the contrary, when teachers assumed the system was created primarily to hold educators accountable and make their jobs impossible, this appeared to lower their system – and personal – value.

Teachers' perceptions of the quality of training and communication over the course of the year improved teachers' perceptions of the purpose of the new system. When the messaging focused on the importance of "teacher growth and development," it increased teachers' value for INVEST. If teachers believed that INVEST was designed to support their growth as professionals, this activated their intrinsic desire to experience success (competence). Conversely, when teachers believed that INVEST devalued their work and served purely an accountability function, this limited their desire to improve practice for their own purposes (autonomy). As one teacher shared, *"it wasn't like I*

wanted to do better because I want to be a better teacher, it was like, omg, if I don't do better, I'm going to be kicked out on my butt" (School 5, Teacher 5). Many teachers reported that they were already motivated to improve their teaching and that evaluation would not have much of an impact on their overall value for teaching in either direction: *"I don't really think an evaluation should be a motivation to be a good teacher. I think that's just part of my job"* (School 5, Teacher 4). Since evaluation was not yet being used for compensation, very few teachers discussed the "utility value" or long-term benefit of performing well on the new evaluation system.

Accuracy and fairness of the measures. During the pilot year of INVEST, teachers' expectancy was influenced by their perceptions of the accuracy and fairness of the measures and evaluation process. Table 7-4 Model 3 showed that the Observation Factor (capturing teachers' perceptions of the accuracy of the observation criteria and process) was a significant predictor of system expectancy, $B = .12$, $t(908) = 19.64$, $p < .05$; however, this predictor was no longer significant when including school effects. Table 7-4 Model 3 also demonstrated that the Student Growth Factor (capturing teachers' perceptions of the accuracy of the growth measure) was an even more influential predictor of system expectancy, $B = .37$, $t(908) = 19.64$, $p < .05$ and that this factor was robust even with the inclusion of school effects. Indeed, teachers' perceptions of the accuracy and fairness of the measures were critical to their initial expectancy on the new system.

Most teachers reported that the Danielson Framework captured a comprehensive and accurate picture of their performance. In particular, they appreciated the specificity of

the Danielson Framework because it provided them with very concrete steps to take to improve practice, which increased their level of system expectancy. However, as discussed in Chapter 4, teachers were very concerned about the achievability of Level 4 performance on the Danielson Framework. As one teacher shared, "*Level four is like ideal. It's like what I've heard from teachers and other administrators is that it's harder to get to level four in INVEST than it is to get that same level with whatever else we were using before"* (School 5, Teacher 1). These concerns over the unattainability of Level 4 performance were often rooted in perceptions of unfairness. One novice teacher shared that she wanted to a Level 4 teacher, but felt it was not feasible as a new teacher: *"To score a four you almost have to be in a leadership position. You have to be a lead teacher…I know that as a first year teacher I'm probably not going to get a four"* (School 2, Teacher 4). Even if teachers believed they were capable of achieving Level 4 performance, many questioned whether the process would provide them with the support they needed to take their practice to the new level. As one teacher remarked:

> I think using those if you could be in the classroom or video it all the time it would be accurate but because you're still only being looked at for those 45 minutes or those 15 minute walkthroughs, you never know when they're going to do that walkthrough. So if it's a Monday and you're introducing a new topic, which to me makes most sense to do as a whole group, then your INVEST is going to reflect poorly because it's not student centered… But then maybe if they'd come in on Friday when we've been doing this, it's not a new skill anymore and now they can do activities in small groups or centers. It kind of depends on the week, and it depends on the topic (School 2,Teacher 5).

Many teachers believed in their ability to reach Level 4 performance, yet questioned whether the observation process would fairly assess those abilities. Teachers expressed concern about how the timing of observations contributed to their accuracy and were

particularly concerned about the walkthroughs (which were unannounced). Since teachers had no way of anticipating when they would occur, this resulted in lowered expectancy that the observation would fairly capture their capabilities.

Teachers also consistently shared concerns with the SGP measure though these concerns remained hypothetical (as SGPs had yet to be fully implemented), rather than based on experience. Most concerns about the accuracy of SGPs were rooted in the lack of student and parent accountability. Teachers raised several specific issues which lowered expectancy on the measure:

- ***Special education students.*** *How is the system fair if SPED students will not be treated as a different sub-group but compared with the general student population?*
- ***Severe behavior problems or other issues.*** *How will the system account for situations where there are a few students in the class who have severe behavior or home-life challenges and disrupt the entire class?*
- ***Attendance.*** *How is the system fair if it doesn't factor in student attendance?*

The principal at School 5 anticipated SGPs being a major concern for teacher expectancy: *"They're not going to like it. Until we can get it figured out for all the particulars, like the teachers who work with special education students, our co-teachers, the teachers who work with bump up students…you're always going to have teachers who do not think it's fair."* Since many teachers perceived student progress on standardized tests to be

influenced by factors outside of their control, this suggests that those teachers will have lower system expectancy.

When INVEST measures aligned with teachers' own definitions of effective teaching, this reinforced their intrinsic value and built overall motivation. For Danielson, this meant valuing student-led classrooms, while for SGPs, it required viewing students' progress on standardized assessments as a meaningful measure of teacher performance. Though many teachers questioned the feasibility of creating student-centered classrooms, they generally valued student-led instruction and believed it was a worthy (though perhaps, unattainable) goal. Though there was considerable variation in teachers' perceptions of the validity of SGPs, more teachers expressed concerns over whether the measure would actually capture valued outcomes. Some teachers pointed out that students would get nervous on the day of the test and not demonstrate their capabilities: *"Just the testing. I don't know. In ways I do feel like a test measures student growth, but in a way I don't. Like it should be more like what they can do in class and how they're thinking. Because some of them when they get to a test, they just clam up. It's not really accurate"* (School 3, Teacher 1). Others went further and questioned whether test performance was actually the goal of education (as opposed to moral/civic development or higher level critical thinking) regardless of whether students performed their best on assessment day.

Support. In addition to their level of understanding and perceptions of the accuracy of the measures, teachers were also influenced by the perceived quality of support associated with the evaluation process. Table 7-4 Model 3 showed that the

Support Factor (capturing teachers' perceptions of the quality of feedback and support) was a significant predictor of system expectancy, $B = .28$, $t(908) = 19.64$, $p < .05$ and that this was robust even with the inclusion of school effects, $B = .25$, $t(874) = 12.90$, $p < .01$. As discussed in Chapter 6, when teachers received quality support to help them reach standards, they were more likely to report believing in their ability to improve on the new system (i.e., higher system expectancy).

System Characteristics and Teacher Effectiveness

As discussed in Chapter 6, teachers' ability to improve their performance on the new system depended on the principal's ability to create systems of support and reflection. Indeed, targeted support is essential to the theory of deliberate practice. When individuals move from the deliberation to the action phase, they commit to specific goals and translate intentions into changes in practice. Yet increased effort is not sufficient to improve performance. Individuals must work at the edge of their abilities and receive immediate and targeted feedback on specific areas of practice. This section examines how system characteristics influenced teachers' ability to engage in deliberate practice and improve on the Danielson Framework for Teaching (the observation measure). I present the same set of analyses discussed in the previous chapters on individual and school characteristics, this time with a focus on system characteristics. Qualitative data confirmed that several system features were particularly important in influencing teachers' ability to improve their effectiveness on INVEST.

Goal-Setting. As theory demonstrates, setting and achieving interim goals increases motivation and in turn, builds interest (or value) in the task itself. Table 7-3

demonstrated that teachers' perceptions of the quality of the goal-setting process were positively related to whether the evaluation system led them to implement changes in practice, $r = .54$, $p < .001$. With the use of the Teachscape technology, one advantage of the feedback principals provided was that it was very detailed so teachers could set goals related to particular components of the Danielson Framework. When teachers were able to successfully implement these incremental changes, the evaluation process enhanced their feelings of competence and helped them maintain commitment.

Table 7-6

Regression Analysis Predicting Teacher Effectiveness on Danielson Observation Measure

	Without School Effects			With School Effects		
Measure	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.09(0.04)*	0.03(0.05)	0.03(0.04)	0.12(0.04)**	0.07(0.05)	0.06(0.04)
Overall Grit	0.01(0.05)	0.00(0.04)	0.01(0.04)	0.00(0.04)	-0.01(0.04)	-0.00(0.04)
Conscientiousness	0.15(0.04)***	0.15(0.04)**	0.14(0.04)***	0.15(0.04)***	0.14(0.04)***	0.13(0.04)**
Extraversion	-0.02(0.04)	-0.02(0.03)	-0.02(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)
Agreeableness	-0.10(0.03)**	-0.10(0.03)**	-0.10(0.03)**	-0.10(0.03)**	-0.09(0.03)**	-0.09(0.03)**
Emotional Stability	-0.12(0.03)**	0.12(0.03)***	0.11(0.03)***	-0.09(0.03)**	-0.09(0.03)**	-0.09(0.03)**
Openness	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.01(0.03)	-0.02(0.03)
School Characteristics						
Quality of Administration		0.11(0.04)*	0.08(0.05)		0.12(0.04)**	0.09(0.04)*
Positive Support		0.07(0.03)*	0.08(0.04)*		0.07(0.03)	0.07(0.04)
Level Control		0.09(0.03)**	0.07(0.03)*		0.10(0.03)**	0.08(0.03)*
Professional Community		-0.14(0.04)***	-0.12(0.04)**		-0.19(0.04)***	-0.16(0.05)**
System Characteristics						
Impact Factor			-0.07(0.08)			-0.07(0.08)
Observation Factor			0.11(0.06)*			0.09(0.06)*
Understanding Factor			0.13(0.06)*			0.18(0.07)*
Support Factor			-0.16(0.10)*			-0.15(0.09)*
Student Growth Factor			-0.03(0.08)			-0.02(0.08)
Goal Factor			0.03(0.06)			-0.01(0.07)

Note: All continuous variables have been standardized. *$p < .05$. **$p < .01$. ***$p < .001$

One teacher described how she found the specific nature of the process especially valuable for her own goal-setting:

> My post conference right here, you can see I did lots of reflection to see exactly what it is I could have done a little bit differently. Just looking at the question – comment on your classroom procedures – all of these components right here. I have to go back and see what is 2C, what is 2D and what is 2E and I need to at least touch on all of those right here…. The level of detail is very important because it tells me what to do (School 1, Teacher 6).

In theory, the specificity of the Teachscape system meant that teachers could isolate specific areas and work strategically to develop those strengths; however, in practice, teachers were often too overwhelmed by the quantity of the feedback to implement changes. While the feedback from PDAS had been confined to one sheet of paper (with check boxes and a few notes), principals had been trained through INVEST to "script" lessons and provide a detailed account of student and teacher interactions. As a result, the feedback obtained through INVEST could be as long as five to ten pages and to access it, teachers had to log-in to the online Teachscape system. Several principals shared that the technology was a challenge for many of their teachers: "*They're very intimidated by it. They couldn't find their information half the time. They didn't know how to use the tool. And they struggled with understanding, for instance, when you score and you have the statements and then the component score, what does all the information actually mean?"* (School 4, Principal).

Understanding. Though the online system had the potential to provide quality feedback, teachers reported needing a better picture of what Level 4 looked like in

practice. As discussed in Chapter 6, principals who developed strong systems of support were able to build teachers' understanding and confidence in the accuracy and fairness of the Danielson Framework. As demonstrated in Table 7-6 Model 6, when individual, school, and system characteristics were simultaneously entered into the regression analysis (and school effects were included), teachers' level of understanding was the strongest predictor of Danielson performance, $B = .18$, $t(847) = 4.90$, $p < .05$, followed by their perceptions of the accuracy of the Danielson observation measure, $B = .11$, $p < .05$. When system characteristics were included in Model 3, the quality of administration was no longer a significant predictor, suggesting that administrators influenced their teachers' performance on INVEST through their ability to build teachers' understanding and confidence in the new system.

Support. Theory indicates that teachers will only be able to sustain their work towards goals when they not only understand the performance criteria but also receive consistent information about how their performance relates to a specific set of standards. Though the Support Factor was not a significant predictor in the regression analysis in Table 7-6 (likely because of its sizeable correlation with other system features), it was the most correlated with reported changes in practice of any of the system features in Table 7-3, $r = .55$, $p < .001$. Additionally, in interviews, teachers reported that the immediacy of feedback was essential in helping them improve their practice. At the beginning of the year, principals experienced a fairly steep learning curve with the technology, which meant that feedback was often not received within the expected one week time frame. For principals who were technologically savvy, the immediacy of the feedback loop

accelerated over the course of the year, and when feedback was received in a timely manner, it helped teachers make incremental adjustments in practice. One new teacher shared more about how this feedback process benefited her development:

> He told me specifically what he wanted to see each time.…He would send me the feedback that day and when we met…he just had a printout of his observation and he just basically walked through it and told me what he thought about it and how I should improve and if I had anything to say to him about his observation and if he missed anything before he walked in and stuff like that. It had a positive impact because I feel that, like I said, at the beginning it's very specific. I like that about the INVEST program. It is very specific and you get it right away. It doesn't leave you wondering where can you improve or where you're doing good or what things are not so good. I like that and I like how everything stays online so you can go back to it.

Though some principals improved the immediacy and consistency of their feedback, others struggled to complete observations in a timely fashion. When teachers did not receive quality feedback on their performance, they were frustrated and did not see the value in the new system. One particularly disgruntled first year teacher noted: *"When we first met [referring to our initial interview], I didn't have much of an idea. I hadn't been observed…I'm like, OK, I thought the idea was to get feedback, especially as a first year teacher. What am I doing wrong? What am I doing right? I'd like to modify what I'm doing. If not, then what's the point of the new system"* (School 6, Teacher 3).

System Characteristics and Teacher Retention

Teachers must sustain improvements in practice by avoiding becoming burnt out by their work. Burnout stems from teachers' motivational responses and is characterized by exhaustion, cynicism about one's value or impact, and frustration over lack of competence. As Chapter 6 made clear, teachers had varied perceptions of how INVEST might impact teacher burnout and turnover across schools, and these diverse opinions

appeared to be based on differing perceptions of system features. To evaluate which of these features were most influential in predicting teachers' turnover, I present the same analysis discussed in Chapters 5 and 6 with a focus on system-level characteristics.

Table 7-7 presents the results when these variables were simultaneously entered into a binary logistic regression model while controlling for individual and school level factors. As demonstrated in Model 3, perceptions of the perceived impact of INVEST emerged as a significant predictor of teachers' retention, and teachers who were one

Table 7-7

Binary Logistic Regression Table Predicting Turnover

	Without School Effects			With School Effects		
Measure	Model 1: *Individual Variation*	Model 2: *Individual and School Variation*	Model 3: *Individual, School, and System Variation*	Model 4: *Individual Variation*	Model 5: *Individual and School Variation*	Model 6: *Individual, School, and System Variation*
Individual Characteristics						
Teaching Grit	0.79(0.05)**	0.88(0.06)	0.83(0.10)	0.79(0.06)**	0.85(0.07)*	0.82(0.11)
Overall Grit	0.92(0.08)	0.94(0.08)	0.94(0.11)	0.92(0.08)	0.93(0.08)	0.91(0.13)
Conscientiousness	1.18(0.16)	1.17(0.10)	1.19(0.17)	1.18(0.12)	1.17(0.11)	1.25(0.19)
Extraversion	0.92(0.06)	0.92(0.05)	0.99(0.11)	0.90(0.05)	0.90(0.05)	0.99(0.10)
Agreeableness	1.01(0.08)	1.04(0.08)	1.11(0.12)	1.01(0.08)	1.03(0.08)	1.10(0.13)
Emotional Stability	1.04(0.09)	1.05(0.07)	0.96(0.12)	1.09(0.08)	1.10(0.12)	1.01(0.13)
Openness	1.04(0.09)	1.04(0.09)	1.09(0.15)	1.05(0.09)	1.05(0.09)	1.11(0.16)
School Characteristics						
Quality of Administration		0.82(0.06)**	0.68(0.07)**		0.81(0.06)**	0.70(0.08)**
Positive Support		0.98(0.08)	0.97(0.10)		0.94(0.08)	0.86(0.09)
Level of Control		0.89(0.06)	0.91(0.12)		0.90(0.06)	1.00(0.15)
Professional Community		1.04(0.07)	1.11(0.12)		1.15(0.08)	1.22(0.32)
System Characteristics						
Impact Factor			0.41(0.15)**			0.40(0.14)*
Observation Factor			0.98(0.15)			1.02(0.17)
Understanding Factor			0.85 (0.19)			0.87(0.20)
Support Factor			1.32(0.32)			1.22(0.32)
Student Growth Factor			3.13(1.07)**			3.38(1.30)**
Goal Factor			0.92(0.26)			0.93(0.27)

Note: All continuous variables have been standardized. Only 873 teachers had data on all variables included in each analysis, so this was the final analytic sample used for all models. Teachers who left the district were slightly less likely to complete end of year surveys, which explains why the aggregate turnover rates are lower than those reported in Chapter 4. *$p < .05$. **$p < .01$. ***$p < .001$

standard deviation higher in perceptions of INVEST's potential impact were 59% less likely to leave teaching in the district, $OR = .41$, $p < .05$. Model 6 indicated that this finding was robust to the inclusion of school effects.

Impact (Aligning with Teachers' Values). INVEST had the potential to comport with teachers' desire to have a positive impact on students. For teachers who saw the primary purpose of INVEST as encouraging their professional growth, they stayed energized by the system. As one invested teacher shared, *"I have never been asked to do this type of reflection before. This is making me a better teacher and keeping me energized to improve"* (School 1, Teacher 3). For this teacher (and others who recognized INVEST's ability to support their development), the new system helped sustain their level of engagement in teaching. To the contrary, when teachers saw the primary purpose of INVEST as "holding teachers accountable," this did not align with their intrinsic values. One teacher shared, "*Yes, I am staying next year but not over the long term. Important to work with the kids, but we don't have enough time to work with the kids. Less testing. Less paperwork. I feel like I'm a secretary*"(School 2, Teacher 2). As this quote suggests, teachers struggling to adapt to – and find meaning in – the additional workload from INVEST were more likely to report wanting to leave the district.

This value conflict was particularly problematic for teachers who associated INVEST with more "unnecessary paperwork" and "testing." While these teachers expressed frustration with the additional workload, their concerns primarily arose from the fact that they did not value the specific type of work principals asked them to engage in under the new system. As one burnt-out teacher shared, *"Just the testing. I don't know.*

In ways I do feel like a test measures student growth, but in a way, I don't. Like it should be more about what they can do in class and how they're thinking. That's what matters to me" (School 3, Teacher 2). Teachers like the one above, felt that "teaching" had become too focused on "testing" which was not why they had joined the profession.

Summary

In sum, this chapter reinforced the initial descriptive findings, presented in Chapter 3, which suggested that teachers' attitudes influenced their experiences with the new system. In fact, perceptions of system features were more highly correlated with teachers' motivation on the system and reported changes in practice than either individual or school characteristics. Even when controlling for these factors, teachers' system perceptions explained considerable variation in motivation, effectiveness, and retention.

Though teachers' perceptions of system characteristics were highly correlated with each other, several emerged as particularly influential. Specifically, teachers' level of understanding of the new system seemed to positively affect their motivation and performance. Indeed, teachers who believed INVEST was designed to support their professional growth as educators and understood the system's expectations were more empowered to take ownership over their practice and reached higher levels of performance on the new system. Teachers' perception of the accuracy of the evaluation measures (both the Danielson Framework and Student Growth Percentiles) was associated with their initial motivation. Teachers' system value was more influenced by their perceptions of the Danielson Framework, which is consistent with the fact that this measure was the focus during the pilot year. However, teachers' system expectancy was

more influenced by their perceptions of the Student Growth factor. This suggests that teachers' belief in their ability to improve performance on INVEST was influenced by whether they had confidence that they could impact students' progress. Finally, the level of support and quality of feedback teachers received was a significant predictor of their initial motivation to improve performance over the course of the year.

The second part of this dissertation has demonstrated that teachers' motivational responses varied considerably depending on individual, school, and system characteristics. Indeed, the overall improvement process (from motivation to volition to commitment) depended on features of individual teachers' personality, conditions within the school, and how teachers' reacted to specific system attributes. Though some characteristics consistently served as positive predictors across all outcomes (i.e., grit in teaching, administration, level of system understanding), other individual characteristics (e.g., openness), school conditions (e.g., level of control) and system features (e.g., accuracy of measures) influenced some outcomes but not others. The final chapter will synthesize these findings in the context of expectancy-value theory to build a stronger understanding of how new teacher evaluation systems influence teachers' motivational responses.

CHAPTER 8: DISCUSSION AND IMPLICATIONS

The purpose of this study is to build a better understanding of the factors influencing teachers' responses to new performance management policies and explore how these responses translate into teacher effectiveness and retention. To investigate these issues, I used mixed methods analysis to examine the impact and implementation of INVEST, a new teacher performance management system which was piloted in Aldine ISD during the 2012-2013 school year. My quantitative analysis captured broad-based results through a survey of the population of teachers in the district and an examination of administrative records. To supplement this analysis, my qualitative research provided a more in depth account of how a subset of individuals experienced policy implementation across different contexts. My findings highlighted that during the pilot year, INVEST had a negative impact on teachers' personal expectancy, but did not have a statistically significant impact on either teacher performance or retention. However, there was considerable variation across all three outcome measures, which was influenced by teachers' perceptions of system features, their individual personality characteristics, as well as elements of school context.

Although a growing body of research has begun to examine the impact of recent evaluation systems, we have very limited knowledge of how individual, school, and system characteristics influence teachers' motivation and improvement process. This is the first study to systematically examine each of these factors and situate findings within a motivational framework. It is my hope that education stakeholders will use these

findings as a means to better understand how teacher evaluation systems impact teachers' motivation, performance, and retention, and how various characteristics influence the implementation process. In this concluding chapter, I return to the literature to compare my findings from this study to the nascent body of research on teachers' responses to performance management systems, as well as the conceptual framework presented on motivational theory, depicted in Figure 8-1 below. I then state the limitations of this analysis and explore broader implications for researchers and policymakers.

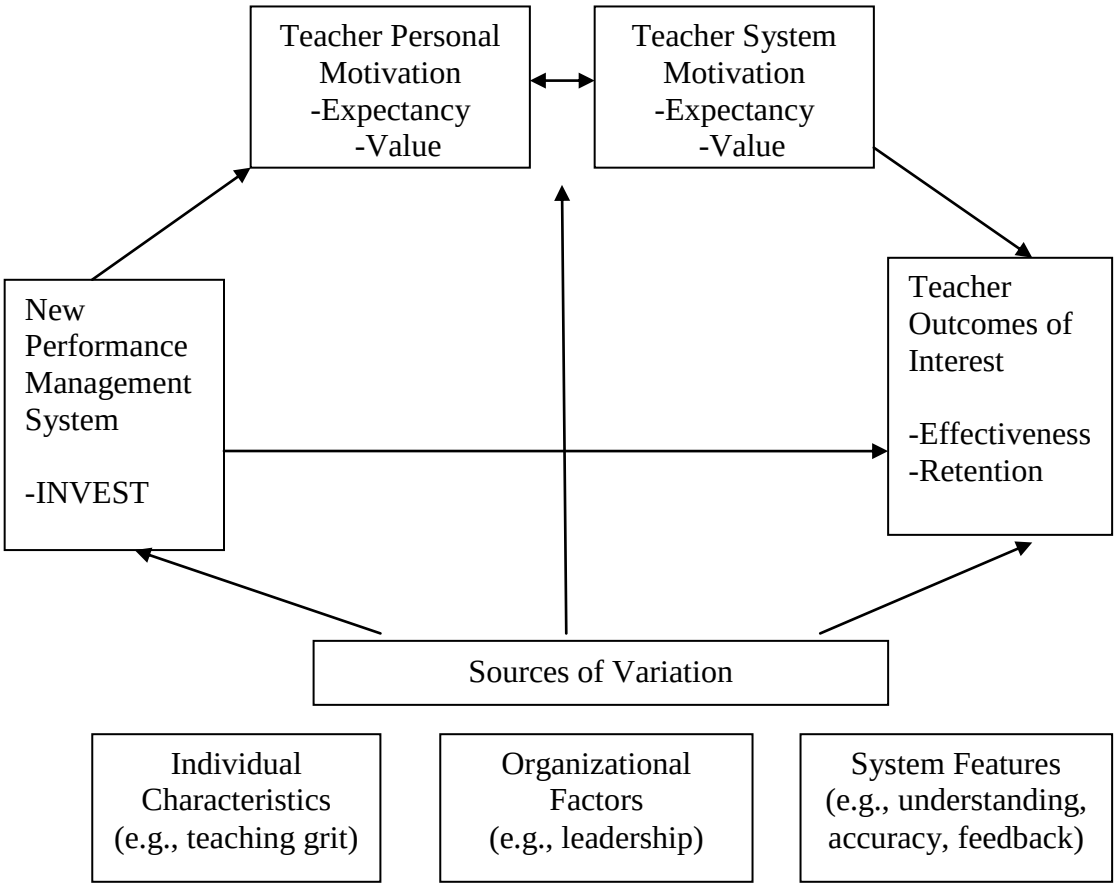

Figure 8-1. Motivational framework based on analysis

Discussion

Understanding Teachers' Responses within Motivational Theory

Teachers' Personal Motivation for Teaching

In the existing psychological literature, expectancy-value theory links motivational choices to two sets of beliefs – the belief an individual has in their own abilities (expectancy) and the value they associate with various tasks (value). Research has demonstrated that more efficacious teachers have a greater influence on student learning than teachers with lower efficacy (Goddard, Hoy, & Woolfolk Hoy, 2004; Tschannen-Moran,Woolfolk Hoy, & Hoy, 1998). In this analysis, teachers' expectancy was shown to influence their performance on the Danielson Framework for Teaching, but this did not translate into their impact on Student Growth Percentiles during the pilot year. When teachers believed in their abilities as educators, they set ambitious goals for their students' performance; conversely, teachers with lower expectancy expressed concerns about how they could get through to the most difficult students who did not come to school motivated or ready to learn.

To be motivated, individuals must not only believe they can make changes in their behavior but also value their work. Value can be intrinsically motivated, based on the level of enjoyment teachers get from a specific task or the extent to which a task is consistent with their self-image (attainment value) or they can be extrinsically motivated, based on perceived utility value or pressure from external sources. Teachers in this

analysis reported very high levels of value for teaching, which was primarily driven by intrinsic motivation for their work. The majority of teachers shared that they entered the teaching profession because they wanted to make a difference in the lives of their students. This is consistent with the first element of self-determination theory, which demonstrates that individuals have a basic desire to experience *competence* in their work. Additionally, many teachers reported that they appreciated being recognized by their administrators as a source of attainment value or validation of their hard work, which aligns with the second element of self-determination theory, *relatedness,* or a desire to positively connect with others. Very few teachers mentioned utility value (or some external benefit) associated with their teaching. When asked about the desirability of performance-based pay, most teachers shared that while they would appreciate extra money, it was not what motivated them. What they appeared to value more was being given ownership over their practice, which is consistent with the third psychological need discussed in self-determination theory, *autonomy.*

From Personal to System Motivation

Even if teachers were motivated by their work, this study demonstrated that teachers' personal motivation for teaching did not necessarily translate into their motivation to perform well on the new system, which is reflected in Figure 8-1. When the new evaluation system rolled out, teachers received information and made judgments about their belief in their abilities to meet system standards and determined the value they placed on performing well on the new system. This system level motivation was influenced by a number of key system features.

Level of Understanding and Perceived Purpose. Teachers' personal motivation was only activated into system motivation if teachers understood how to improve their performance under the new system and believed it was designed to support their professional growth. Previous research has documented that teachers often react negatively to policies because they do not understand how the policy is designed to operate or they perceive performance metrics to be unattainable. In an analysis of Florida's Merit Award Program (MAP), 61% of teachers reported having little understanding of how MAP measured high quality teaching which contributed to the fact that only 35% believed it was fair for teachers to receive pay based on value-added results (Jacob & Springer, 2007). Teachers' overall level of understanding of the new INVEST system was similarly quite low, with only 53% agreeing or strongly agreeing that they had a solid initial understanding of the system's expectations at the beginning of the year. This analysis also demonstrated that teachers who believed that INVEST was designed to support professional growth were more optimistic about their ability to improve outcomes on the new system. When the messaging associated with INVEST focused on "teacher development" as opposed to "teacher appraisal or accountability," this activated teachers' expectancy.

Accuracy and Fairness of the Measures. Expectancy theory makes clear that an individual's motivation will be strengthened when performance goals are clearly defined. This clarity will allow individuals to determine the value they attribute to particular goals and assess how likely they are to achieve them with increased effort (Locke & Latham, 1990). If systems become too complex, they run the risk of resulting in a lack of clarity

and a corresponding decrease in motivation (Heneman, Milanowski, & Kimball, 2007). As such, teachers' perceptions of accuracy and fairness of the measures are essential to building teachers' system expectancy**.** This study corroborated prior research that performance management systems are only motivating when teachers believe that the criteria used to gauge effectiveness are fair (Johnson & Papay, 2009). Though most teachers believed the Danielson Framework was both accurate and comprehensive, they questioned how fair it was to expect teachers to create student-driven classrooms (the hallmark of the top performance level, Level 4, on the new system). Teachers also questioned whether it was fair to hold them accountable for students' progress (through the SGP measure) when there was a lack of student and parent accountability. Though teachers generally had better perceptions of the Danielson Framework than SGPs, they did not consistently prefer one over the other and expressed concerns with both performance measures.

Given the complexity of teaching, it has historically been extremely challenging to develop measures for evaluating teacher practice as part of performance management systems. Often referred to as the "nature of teaching" hypothesis, the fundamental challenge in determining teacher quality has always been how to clearly define outcomes and separate the impact of the teacher from other influences on student learning (Podursky & Springer, 2006). As discussed in Chapter 1, a considerable body of existing research has focused on validating performance measures. However, less research has investigated teachers' value for these metrics. This analysis revealed that teachers' definition of validity and reliability differed from those of statisticians. Without advanced

statistical knowledge, teachers did not fully understand how growth models attempted to estimate teacher effects or how inter-rater reliability was calculated. Instead, they cared about face validity and were motivated when measures aligned with their individual values.

Feasibility of the Expectations. For the Danielson Framework, it was not the measure itself that concerned some teachers; indeed, they generally believed that it captured a comprehensive picture of their performance. Instead, they raised expectancy concerns about whether achieving top performance on the measure was a feasible expectation, which in turn, appeared to influence the value they placed on their system performance. For SGPs, many teachers were resistant to the idea of including student growth as part of their evaluation regardless of how they performed on the measure. These teachers questioned whether standardized tests accurately capture student learning and believed the goal of education should be focused on higher order thinking skills and performance throughout the year, rather than reduced to performance on a single day of testing. In a prior study of teachers in Washington State, only 17% of those surveyed were in favor of incentive pay based on test score gains (Goldhaber et.al, 2007). Though the question was not phrased in terms of incentive pay, the results were similar in Aldine ISD – only 30% of teachers agreed or strongly agreed that SGPs were a fair and accurate measure of their performance (compared to 37% for the Danielson Framework).

High Level of Attainment Value. Despite concerns over performance measures, the majority of teachers still expressed a strong desire to reach the top level of the evaluation system (82% of teachers reported valuing performing well on the new

system). This is consistent with other research in Texas which examined teachers' motivation on a previous performance management system, TEEG (Springer et al., 2008). For many teachers, this was a matter of pride; indeed, achieving highly effective status was necessary to maintain their self-image. Teachers' frustration with the new system often resulted from being labeled as "proficient" or "effective" which indicated a high level of initial attainment value. This finding suggests that the variation in value was in part driven by teachers' responses to the labeling of their effectiveness on the system itself.

From System Motivation to Effectiveness

Several recent studies have found that teacher evaluation can lead to increased student learning. For example, in Cincinnati, a student instructed by a teacher after participation in the new evaluation system was projected to score about 11% of a standard deviation or 4.5 percentile points (for a median student) higher. Though the authors were not able to identify the mechanisms driving these improvements, they speculated about several possible factors based on the system's design – the usefulness of feedback, the self-reflective process, and the quality of conversations between teachers and administrators about practice (Taylor & Tyler, 2011). Though the analysis of INVEST did not reveal a similarly positive impact on teacher effectiveness overall, similar factors did emerge as particularly influential in the improvement process. Indeed, teachers in schools with high implementation fidelity received more targeted feedback and ongoing support, and further, were empowered to take ownership over their own

goal-setting process. As a result, they reported being more likely to implement changes in their practice under the new system.

Implementation Fidelity. Many principals experienced a steep learning curve during the first few months of INVEST implementation and struggled to provide teachers with timely and targeted feedback that could be used to improve performance. Additionally, while some principals empowered teachers to self-reflect and guide the dialogue about their practice, others adhered to a top-down structure and overly managed the evaluation process. Much like teachers struggled to create student-driven classrooms (or reach Level 4 performance), many principals were similarly unable to invest teachers in the improvement process. As such, only 45% of teachers in pilot schools agreed or strongly agreed that the new system led them to improve their practice.

The Feedback Loop. In addition to receiving feedback about how their performance relates to a specific set of standards, individuals need to use this information to set goals for future performance (Bandura, 1982; Bandura & Schunk, 1981) and engage in deliberate practice (Ericsson, 1993). This analysis provided evidence in support of this theory. Indeed, teachers' level of engagement in the goal-setting process and their perceptions of the quality of feedback were both associated with whether they reported the evaluation system led to changes in practice. Since the Danielson Framework was very specific, it allowed teachers to effectively set interim goals over the year. For teachers who actively participated in this goal-setting process, achieving interim goals led to increased satisfaction, which, in turn built value for the task itself.

From System Motivation to Retention

To sustain commitment to their work over time, teachers must avoid experiencing burnout. In Aldine ISD, teachers' ability to stay engaged with INVEST was influenced by the implementation of the new system, in particular, the extent to which teachers believed the new system was designed to support their professional growth and empowered them as professionals. It is important to note that not all retention is desirable. Indeed, one of the goals of INVEST (as is the case with most other performance management systems) was to increase the attrition rate of ineffective teachers. The study conducted of Washington, D.C.'s IMPACT system found that dismissal threats increased the voluntary attrition of low-performing teachers by .27 of a standard deviation (Dee & Wycoff, 2013). This analysis also discovered that pilot schools participating in INVEST had an increase in the rate of teacher turnover and that these rates of turnover were higher among teachers identified as Needs Improvement or Ineffective under the system. However, since these differences were not statistically significant, they can only be used as suggestive confirmatory evidence.

Perceptions of purpose. Though one of the goals may indeed be to exit underperforming teachers, it is not motivating for teachers if this is how they perceive the system's primary purpose. When teachers believed INVEST was designed to support their development as professionals, they were more likely to stay energized by their work and committed to improving performance. Conversely, when they viewed the system as a mechanism for accountability and dismissal of underperformers, they were more likely to feel threatened and demeaned by INVEST. Indeed, teachers' ability to sustain their motivation and improvements in performance on INVEST appeared to be driven by their

understanding of the system's purpose as well as the value they placed on the additional requirements of the new system. When teachers were not invested in the value of these additional requirements, it just felt like more unnecessary paperwork or the imposition of unfair expectations.

Empowerment. Across the board, the timeline was too rushed at the beginning of the pilot year, which led teachers to feel overwhelmed by new system requirements and had an overall negative impact on expectancy. As the year progressed, certain principals developed systems to better structure implementation and provide teachers with ownership over the improvement process, which led teachers in these schools to increase commitment. In other schools, teachers complained that training was not aligned with other expectations, and though teachers reported having more work than in previous years, they did not feel any more recognized for their contributions. In fact, under INVEST, at the same time teachers were being asked to take on more work, they were simultaneously being told they were no longer at the highest level of the system (which challenged many veterans' sense of competence). Given the many requirements placed upon teachers, the imbalance of demands and resources left a subgroup of teachers feeling burnt out. Conversely, when principals empowered teachers to become agents of their own improvement process, INVEST had the opposite effect and appeared to result in improvements in teacher commitment.

Contributing Factors

As this analysis has made clear, throughout the process, teachers' motivation, effectiveness, and retention were influenced by both individual and school characteristics.

Indeed, there was considerable variation in teachers' responses to the system both within and between schools. This section will explore how teachers' responses and behavior under the new system were influenced by the nature of the worker (individual characteristics) and nature of the working conditions (school characteristics).

Nature of the Worker (Individual Characteristics). To impact motivation, performance management systems need to be congruent with the needs, values, and capabilities of the people they attempt to influence (Lawler, 1983; Vroom, 1964). Yet, most performance management systems treat teachers as a monolithic entity. At the same time policymakers are calling for new systems to differentiate teachers based on their performance level, the assumption seems to be that their motivational responses to specific policies will be consistent.

This study demonstrates that teachers' responses to INVEST varied considerably based on teacher demographics and personality characteristics. Newer teachers were more likely to value performing well on INVEST, though their expectancies were not significantly different from their veteran counterparts as previous research has suggested. Since they received additional feedback and support, reaching higher levels of performance seemed more feasible. Additionally, for many veterans who had always been at the top level of the prior evaluation system, INVEST was a significant adjustment and when they did not reach Level 4 performance, this resulted in increased frustration and lowered expectancy. The exception, of course, was highly effective teachers (on the Danielson Framework) who were considerably more confident in their ability to perform well on the system and in turn, valued that attainment.

In terms of personality characteristics, teachers' grit in teaching was the only variable to influence all three outcomes of interest – motivation, effectiveness (on the Danielson Framework) and retention. This finding corroborates prior research, as well as anecdotal observation, that teaching can be incredibly discouraging work. In a national survey, teachers identified enthusiasm, energy, and effort as critical qualities for classroom success and encouraged only those with a "true sense of calling" to pursue the profession (Farkas, Johnson, & Foleno, 2000). Given challenges associated with teaching, it seems logical that grit would have a positive influence on teacher motivation, performance, and persistence. Consistent with other psychological research, the domain-specific (teaching) grit scale, which was modified slightly from the domain-general (overall) grit scale was more predictive of outcomes.

Though previous research has demonstrated that grit was a significant predictor of teacher performance and retention (Duckworth et al., 2009; Robertson-Kraft & Duckworth, 2014), this is the first study to explore the specific mechanism by which grit translates into teachers' motivation and behavioral change. Of the five personality types described in this analysis, the invested teacher closely captured what we would characterize as a "gritty" teacher. Invested teachers were able to maintain high levels of expectancy and commitment even in the face of significant challenge. Rather than dwell on obstacles outside of their control, they set clear goals for performance and maintained a strong sense of purpose over the course of the year. These goals were not general and aspirational (e.g., "reaching a Level 4"). Rather, they were targeted and specific (e.g., "improve student participation through more strategic questioning techniques") and

embedded in all aspects of their practice. Invested teachers were self-reflective and actively sought out feedback and professional development opportunities to work towards their goals. Instead of becoming discouraged by critical feedback, they relished it as an opportunity for personal growth and stayed optimistic about their abilities to impact student learning over the course of the pilot year.

Particularly in low-income districts, the multiplicity of factors outside a teacher's control (e.g., parental support, available resources, working conditions) obscures the link between hard work and positive student outcomes. According to Lortie, these "endemic uncertainties" associated with teaching have led many teachers to develop a resistance to change because they believe their work environment has never permitted them to demonstrate their effectiveness (Lortie, 1975). In support of Lortie's theory, skeptical teachers, who were practical and analytical individuals who remained fairly neutral toward the system, were by far the most common type of teacher (and much more common than the invested teacher). As veterans, they had seen systems come and go and as a result, were more conservative and focused on their short-term success with students. According to Lortie, since teachers' time is their most precious resource in their quest for psychic reward, teachers resent interruptions and prefer to be left alone (Lortie, 1975). Consistent with this theory, several skeptical teachers became insulted by the end of the year when INVEST did not reinforce their belief in their impact and added what they perceived as unnecessary and additional work to their already overwhelming responsibilities.

In sum, though teachers' gritty nature was internal (as one invested teacher shared "*It's just my personality to be this way")*, this study also provides support for Lortie's analysis of the importance of working environment, as teachers' responses to systems also appeared to be influenced by their school context. Indeed, it is likely not coincidental that several of the invested teachers were clustered within School 1 and that several schools did not have any teachers of this type. The next section will explore how the nature of working conditions influenced teachers' responses to the new evaluation system.

Nature of the Working Conditions (School Characteristics). Both theory and research demonstrate that the effectiveness of a performance management system will ultimately depend on how well it fits within a particular context. Unfortunately, policymakers typically focus most intently on the design features of performance management systems and neglect to address equally important context issues. This study demonstrated that teachers' motivation and subsequent behavioral change were influenced by a variety of enabling conditions, and that the principal was central to developing the climate for effective implementation.

Consistent with prior research (Kelley et al., 2000), principal quality influenced teachers' system motivation, as well as their subsequent performance and retention decisions. Highly effective principals were able to create supportive environments that helped their teachers feel it was possible – and important – to meet new system expectations. However, rather than just implement the new system with fidelity, they used INVEST to empower teachers to reflect on their practice and drive their own

improvement. Rather than present INVEST as another tool for teacher accountability, highly effective principals communicated that the new system was a way to support teachers in their own professional growth and to maximize their ability to impact students. As a result, this activated teachers' value for their system and encouraged them to set meaningful long-term goals.

This analysis demonstrated that if teachers were going to be motivated to improve practice, they needed to be invested in the purpose of the new evaluation system. Highly effective principals clearly communicated that the system was designed to support professional growth and subsequently empowered teachers to take ownership over setting and monitoring progress on their own goals. To support the self-reflective process, they created structured time for teachers to meet with each other to discuss practice in professional learning communities and also utilized the evaluation pre- and post-conferences to develop teacher's self-reflective abilities. As a result, highly effective principals forged strong trusting relationships with their staff, which helped teachers feel more supported in reaching their goals and in turn, ensured they avoided burnout. Conversely, teachers in schools with less effective principals reported that INVEST was top-down and exacerbated an already punitive school culture. Given their frustration with their teaching experience, these teachers were less likely to be motivated to implement changes in practice and stay committed to the profession over the long term.

This study confirms prior research that the distribution of influence and control in schools profoundly affects how they function. Historically, whether districts can successfully adopt reforms has been shown to depend on teacher buy-in and investment

in the process (Hannaway & Rotherham, 2008). In *Who Controls Teachers' Work*, Ingersoll demonstrates that teachers have more control over academic issues (e.g., curriculum) than they do over administrative and policy decisions (e.g., professional development, evaluation); in fact most teachers have little influence over anything but instructional matters (Ingersoll, 2006). If policies are too top-down in their nature, they limit teachers' flexibility to make decisions about what is best for their students. When implementing INVEST, highly effective principals were able to successfully balance the need for accountability (inherent with any teacher evaluation system) with efforts focused on teacher empowerment.

Once teachers were initially motivated, highly effective principals provided targeted feedback and offered opportunities for ongoing support to accelerate their development. In line with the theory of deliberate practice, these principals diagnosed very specific deficiencies in a timely fashion and aligned their feedback with meaningful suggestions for improvement. When teachers did not receive this type of feedback and consistent support, it proved more challenging to develop practice on the new INVEST system. Given the increased expectations associated with implementation, more organized and conscientious principals were better able to effectively manage the demands of the new evaluation system.

Limitations

Several limitations of the current investigation are worth noting. First, given the non-experimental nature of the school selection, third variable confounds pose a potential threat to the internal validity of the impact analysis. The available data made it possible to

demonstrate that the pilot and control schools were equivalent at baseline and to adjust for individual and school level characteristics in the analysis. Thus, the most obvious third-variable confounds were accounted for in the present investigation, though there could have been unobservable variables impacting outcomes. It is important, therefore, that future research continue to investigate the impact of new teacher performance management systems with more rigorous experimental designs.

Second, although the Student Growth Percentile measure captured information on teachers' impact on student progress (rather than absolute achievement), it had several shortcomings. First, the ratings were not adjusted for student or school characteristics, which some research demonstrates has the potential to influence student academic gains. Second, because the ratings were only based on one year of data, they presented a limited picture of teachers' impact on student progress. As history makes clear, defining quality teaching is an incredibly challenging task, and scholars have contested the validity and reliability of various performance metrics. From a motivational perspective, it is perhaps even more important to note that teachers in pilot schools did not receive their SGP ratings during the pilot year, and thus were not able to use them to understand their performance or improve their practice. As a result, the SGP measure may not have been the best mechanism for capturing information on teachers' improvement over the course of the pilot year, and future research should use multiple measures (including observational measures like the Danielson Framework for Teaching) to assess the impact of new policies.

Additionally, this investigation only examined the impact of the new teacher performance management system during the pilot year of the initiative. As the descriptive analysis demonstrated, many principals faced considerable challenges with INVEST implementation, and teachers noted that it would take time to adjust to the new expectations. As a result, it is certainly possible that with additional support – or just additional time – the results of an impact analysis would be different in subsequent years. However, it is unlikely that the drivers influencing variation in responses at the system, school, and teacher level will change considerably. Indeed, qualitative research on INVEST has carried into the second year of implementation and we have discovered similar trends in teachers' responses both across and within schools.

Finally, it is important to note that the external validity of these findings is limited by the nature of the sample. Since I studied the implementation of a new teacher evaluation system in a relatively large urban district in a non-bargaining state, these findings may not generalize to teachers in different types of districts. Research has demonstrated that rates of teacher turnover are much higher in urban contexts, suggesting that the experience of urban teachers differs in important ways from those in suburban or rural districts. In addition, scholars have historically documented that performance management policies encounter intense resistance from teachers unions. Since Texas is not a bargaining state, Aldine ISD did not have to negotiate the design or implementation of the new teacher evaluation system. In districts with strong unions, we might expect that the collective bargaining process would influence the policy design, as well as how teachers responded to the new initiative during implementation.

Implications

Despite these limitations, the current analysis makes a significant contribution toward understanding how system, individual, and school characteristics influence teachers' motivation, performance, and retention under new teacher performance management systems. The growing interest in these systems offers an important opportunity for further policy development and research as these initiatives proliferate across the country. This section provides several important recommendations for policymakers and researchers.

Future Policy

Focus on the perceived accuracy and fairness of performance measures. During the pilot year of INVEST, teachers' expectancy and value were both influenced by their perceptions of the accuracy and fairness of the measures and evaluation process. When INVEST measures aligned with teachers' own definitions of effective teaching, this reinforced their intrinsic value and contributed to overall motivation. Teachers were also motivated by measures when they were clearly articulated and provided a pathway for improvement. Though they appreciated the specificity of the Danielson Framework, many teachers were concerned about the achievability of Level 4 performance and insulted by only being able to achieve "proficient" or "effective" status. Teachers concerns with the fairness of SGPs were rooted in how the measure would capture factors they perceived were outside of their control (e.g., student or parent accountability). The challenge for policymakers will be to employ measures that produce results teachers view as accurate and provide sufficient training to help them interpret and utilize data to

improve their instructional practice. Additionally, policymakers should carefully consider how to label performance levels, so as not to be demotivating for teachers. To accomplish this, initiatives will likely need to include multiple measures that provide a more comprehensive picture of teacher effectiveness and focus intently on communication efforts during the early months of implementation.

Embed evaluation as part of broader development effort and message the system's purpose. Improved evaluation is not a panacea; indeed, it will not lead to increased teacher effectiveness unless it is accompanied by conditions that build teacher expectancy and value. Scholars contend that today's attempts at performance management are not narrowly focused but instead include other initiatives (e.g., development, compensation) that directly influence the objective of increasing teacher quality (Odden & Wallace, 2004). This study supports prior research which has shown that new teacher evaluation systems will be more effective if they spur teachers' investment in their own long-term development (Taylor & Tyler, 2011). As these systems proliferate, it is essential that policymakers do not view evaluation as an isolated strategy, but instead, as part of a comprehensive system designed to recognize and improve instructional expertise. Most importantly, they must communicate this purpose to teachers. When teachers view new systems as designed to support their professional growth, as opposed to hold them accountable for performance, they are more likely to be motivated to sustain changes in behavior over time.

Develop principals to support their teachers to engage in deliberate practice. Motivational theory demonstrates that goals will be more motivating when individuals

not only value the performance criteria but also believe that through hard work, they can successfully improve their own effectiveness on these measures. In other words, evaluation cannot lead to improvements in performance unless teachers also have the capacity to implement necessary changes in their practice. This study demonstrated the important influence of principal leadership on teacher motivation, performance, and retention. To improve performance, teachers need to not only know the expectations, but also be given specific and consistent feedback to implement changes effectively in their own classroom. Policymakers should provide principals with training around coaching and reflective conversations, so they can support teachers to engage in the process of deliberate practice. Additionally, they should support principals to develop the organizational management systems necessary to effectively implement the new policy with fidelity.

Differentiate systems to meet teachers' and schools' varying needs. This analysis revealed that teachers have divergent reactions to performance management systems both within and across schools. Given the influence of individual and school characteristics on outcomes, it is challenging to design a one-size-fits-all approach to performance management. To maximize teachers' motivational responses, policymakers need to gather data to ascertain what is of value to teachers in their district and construct new (and likely, multiple) performance metrics and incentives to adequately reflect these values. To maximize teachers' motivational responses, policymakers should design differentiated tracks that can be better customized to meet the needs of specific subgroups of teachers and provide differentiated support to principals based on the needs of their

campus. For example, since they are new to the classroom, novice teachers will require additional guidance and likely be more open to adopting new practices. In contrast, more experienced teachers (particularly skeptical and insulted teachers) will be less likely to embrace change. Policymakers can increase the likelihood that the new initiative will be responsive to teachers' needs by developing varying communication messages and support structures depending on teachers' and schools' needs.

Involve teachers in the process. This study demonstrated that individuals' motivational responses were influenced by teachers' level of ownership over the evaluation process. Increased involvement built trust and engendered overall commitment to the system. Unfortunately, historical attempts to implement performance management systems have typically neglected to take into account teachers' perspectives and consequently, reforms have not been sustained over time (Tyack & Cuban, 1997). To increase the likelihood of sustainability, policymakers should seek to ensure that teachers play an active role in designing and implementing new performance management initiatives. Indeed, if the goal of these initiatives is to improve the quality of teaching, policymakers must recognize that history has repeatedly shown that those most directly affected by policy must be invested in and empowered by the process of change.

Future Research

Evaluate new teacher performance management policies in the context of motivational theory. As this analysis makes clear, teachers' responses to new performance management systems will vary considerably as a function of differences in system features, individual teacher characteristics and school-based contextual factors.

Unfortunately, most studies do not investigate these types of variation, and when they do, results are not analyzed within the context of motivational theory. In this analysis, I developed a conceptual framework, derived from the literature on motivational theory, to frame how we might expect teachers to respond to new performance management initiatives and then explored how design features, individual differences and school-based organizational factors influenced these motivational reactions. It is my hope that this framework will serve as an analytic tool for future researchers as they seek to understand teachers' responses to new systems.

Use mixed methods analysis. This study indicates that perceptions of new evaluation systems vary considerably based on specific teacher characteristics and contextual factors. Unfortunately, most studies do not take into consideration how individual and school characteristics affect teacher attitudes and subsequently influence motivation and performance. Current performance management policies include a variety of components in their design. To study this complexity, researchers will need to strike the appropriate balance between rigorous quantitative impact research and systematic qualitative analysis that explores how teachers' perceptions influence outcomes. Researchers need to employ mixed methods to develop a better understanding of how new performance management influence subgroups of teachers in different types of contexts.

Validate system measures and outcomes. When exploring teachers' motivation and attitudes towards their work, researchers should pay particularly close attention to how they measure and validate specific constructs. This analysis employed existing self-

report measures for individual and organizational characteristics and also used exploratory factor analysis to develop several new scales for capturing teachers' attitudes toward new evaluation systems. Researchers should validate these scales and determine how they apply across different contexts. Given the demonstrated influence of individual, school, and system characteristics, future research should also examine how these factors moderate the impact of new systems on outcomes of interest. Additionally, given limitations associated with the reliability and validity of various performance measures (e.g., SGPs, Danielson Framework of Teaching), studies should also examine the impact and implementation of new initiatives using multiple measures and then, compare results across outcomes.

Conduct multi-year studies across multiple contexts. This analysis only examines the impact and implementation of one performance management system during its pilot year, which as noted above, limits the external validity of the results. Though there is growing interest in performance management systems nationally, the existing research base is considerably more limited. Researchers should continue to investigate the impact and implementation of these systems across contexts and employ consistent measures (such as the survey metrics used in this analysis) and examine similar subgroups, so that results can be more easily compared across studies. Research should be longitudinal so that we can also build a better understanding of teachers' motivational responses and behavioral changes as they become more accustomed to new systems.

Conclusion

Historical evidence has demonstrated that teacher acceptance will ultimately contribute to the effectiveness and survival of teacher performance management policies. Recent initiatives are also beginning to demonstrate that the success of these new systems can hinge on teachers' reactions to new performance measures and incentives. Yet, the body of research on how teachers respond to new performance management policies remains surprisingly underdeveloped. To complicate matters, the research that does exist reveals mixed results about the validity of performance measures, as well as the impact new systems have on teacher perceptions and student outcomes.

Though many policy briefs purport to inform policymakers of the essential factors to consider when creating performance management initiatives, this is the first analysis to systematically examine the influence of individual, school, and system characteristics on teachers' motivation, performance, and retention on a new teacher evaluation system. Given how widespread implementation of these new systems has become, it is critical that as these initiatives proliferate, they are designed in a way that is amenable to evaluation. Evaluators of current performance management initiatives must pay careful attention to how different design decisions influence teacher expectancy and value and ultimately translate into motivation and behavioral change. Researchers should also consider how these reactions are influenced by individual characteristics of teachers, as well as organizational conditions in schools. In turn, as they implement new systems, policymakers will need to be prepared to revise initiatives as they learn more about teachers' motivational responses to changes in performance management policies.

Appendix

Table 5-1 Supplement

Descriptive Data by Individual Teacher Profile Type (From Survey Data)

Profile Type	% Female	% White	Yrs Exp	INVEST Expectancy	INVEST Value	INVEST Perceptions	SGP	DAN	E	C	O	N	A	G	TG
Invested Teachers	80%	40%	18	4.20	4.70	3.82	20%	60%	3.42	4.58	4.17	1.88	4.50	4.08	4.56
Sponge Teachers	50%	33%	1	3.50	4.25	3.61	17%	0%	3.56	4.22	4.45	2.17	4.56	4.00	3.85
Burnt-out Teachers	83%	83%	3	2.75	4.62	3.04	17%	0%	3.56	3.96	3.67	2.50	4.11	3.67	3.53
Insulted Teachers	83%	33%	6	1.50	3.50	2.44	0%	0%	3.60	4.73	4.13	1.90	4.47	4.53	3.40
Skeptical Teachers	85%	15%	12	3.60	4.10	3.31	31%	30%	3.06	4.06	4.12	2.73	4.33	3.80	3.91

Note: N = 26, as only 72% of the 36 teachers completed both surveys.

Demographics and Performance	*Personality*
% Female and % White = % of teachers who are female and white *Yrs Exp = Average years of experience in teaching* *INVEST Expectancy = Belief in ability to perform well on INVEST (Scale of 1-5)* *INVEST Value = Value placed on performing well on INVEST (Scale of 1-5)* *INVEST Perceptions = Average score of perceptions of INVEST (Scale of 1-5)* *SGP and DAN = % reaching Highly Effective status on Student Growth Percentiles and Danielson*	*E = Extraversion* *C = Conscientiousness* *N = Neuroticism* *A = Agreeableness* *O = Openness* *G = Grit* *TG = Teaching Grit*

Table 6-1 Supplement

Descriptive Data by School Case Study (From Survey Data)

Measure	**Mean (SD)**	**School 1 (102)**	**School 2 (111)**	**School 3 (69)**	**School 4 (63)**	**School 5 (83)**	**School 6 (81)**
Individual Personality Characteristics							
Teaching Grit	3.80	3.94	3.81	3.86	3.88	3.68	3.91
	(0.67)	(0.54)	(0.66)	(0.61)	(0.63)	(0.77)	(0.67)
Grit	3.90	3.89	3.75	3.87	3.91	3.90	3.88
	(0.48)	(0.61)	(0.63)	(0.41)	(0.44)	(0.54)	(0.51)
Conscientiousness	4.33	4.37	4.22	4.39	4.24	4.22	4.31
	(0.50)	(0.54)	(0.58)	(0.52)	(0.46)	(0.54)	(0.56)
Extraversion	3.39	2.89	3.07	3.59	3.24	3.43	3.29
	(0.80)	(0.67)	(0.94)	(0.76)	(0.72)	(0.87)	(0.83)
Agreeableness	4.20	4.33	4.24	4.27	4.11	4.04	4.22
	(0.58)	(0.47)	(0.56)	(0.48)	(0.55)	(0.69)	(0.56)
Neuroticism	2.66	2.56	2.86	2.52	2.71	2.53	2.56
	(0.87)	(0.88)	(0.82)	(0.88)	(0.80)	(0.95)	(0.85)
Openness	4.12	4.19	4.12	4.24	3.99	4.23	4.02
	(0.55)	(0.44)	(0.60)	(0.51)	(0.47)	(0.59)	(0.64)

School Working Conditions							
Quality of	3.80	4.04	3.87	4.12	3.54	4.25	4.20
Administration	(0.88)	(0.67)	(0.84)	(0.72)	(0.93)	(0.54)	(0.68)
Positive support	3.18	3.29	3.19	3.30	3.29	3.38	3.23
	(0.81)	(0.73)	(0.80)	(0.79)	(0.64)	(0.69)	(0.82)
Level of control	3.61	3.43	3.74	3.95	3.71	3.53	3.56
	(0.77)	(0.73)	(0.63)	(0.76)	(0.69)	(0.70)	(0.81)
Professional community	3.48	3.51	3.61	3.74	3.38	3.94	3.75
	(0.79)	(0.70)	(0.75)	(0.63)	(0.76)	(0.55)	(0.71)
Teacher Perceptions of Evaluation							
Evaluation Measures	3.77	3.72	3.39	3.90	3.93	3.95	3.70
	(0.82)	(0.63)	(0.86)	(0.70)	(0.63)	(0.71)	(0.78)
Evaluation Process	3.70	3.75	3.27	3.71	3.48	3.79	3.30
	(0.91)	(0.70)	(1.06)	(0.91)	(0.84)	(0.74)	(0.92)
Frequency	3.83	3.94	3.90	4.01	3.86	3.95	3.51
	(0.96)	(0.63)	(0.93)	(0.86)	(0.82)	(0.81)	(1.00)
Feedback and Growth	3.54	3.39	3.30	3.65	3.43	3.70	3.36
	(0.83)	(0.86)	(1.01)	(0.85)	(0.90)	(0.75)	(0.85)
Teacher Perceptions of INVEST							
Understanding	3.31	3.36	3.20	3.65	3.55	3.25	3.53
	(0.82)	(0.88)	(0.72)	(0.64)	(0.62)	(0.82)	(0.67)
Goal-setting	3.16	3.38	2.95	3.36	3.24	3.15	3.11
	(0.91)	(0.87)	(1.12)	(0.82)	(0.82)	(0.88)	(0.93)

INVEST Measures	3.26	3.45	3.02	3.58	3.57	3.39	3.34
	(0.79)	(0.88)	(0.85)	(0.60)	(0.91)	(0.66)	(0.79)
INVEST Growth and Impact	3.20	3.19	3.07	3.60	3.44	3.35	3.31
	(0.81)	(0.86)	(0.85)	(0.70)	(0.70)	(0.63)	(0.77)
Teacher Outcomes							
Motivation							
Expectancy	3.97	3.78	3.96	3.98	4.10	3.88	3.85
	(0.59)	(0.62)	(0.66)	(0.59)	(0.62)	(0.66)	(0.61)
Value	4.22	4.31	4.27	4.25	4.25	4.05	4.24
	(0.58)	(0.57)	(0.64)	(0.42)	(0.46)	(0.67)	(0.50)
INVEST Expectancy	3.24	3.41	3.15	3.76	3.24	3.44	3.43
	(1.14)	(0.94)	(1.29)	(1.03)	(1.00)	(0.79)	(1.13)
INVEST Value	4.20	4.29	4.17	4.36	4.52	3.95	4.30
	(0.83)	(0.59)	(0.90)	(0.72)	(0.60)	(0.70)	(0.76)
Effectiveness							
Observation (Danielson)	3.19	3.39	3.12	3.22	3.18	3.21	3.24
	(0.33)	(0.36)	(0.42)	(0.35)	(0.26)	(0.31)	(0.31)
% Highly Effective Danielson	14%	54%	13%	27%	29%	18%	14%
Student Growth Percentile	51.61	52.00	60.17	52.10	53.95	55.06	55.50
	(13.07)	(19.44)	(19.15)	(12.33)	(10.42)	(10.20)	(11.81)
Retention							
Teacher Burnout	2.95	2.98	3.10	2.92	2.92	2.99	2.99
	(0.97)	(0.99)	(0.95)	(0.92)	(1.08)	(0.94)	(0.84)

Teacher Turnover Intentions	2.31 (1.01)	2.55 (1.06)	2.46 (0.96)	2.13 (0.90)	2.24 (1.14)	2.39 (0.97)	2.39 (0.86)
Teacher Turnover (from the District)	13%	20%	13%	9%	14%	11%	22%
Sample Size		17	52	44	21	53	46

N = 2662. For pilot measures, *N* = 1097 and for Student Growth Percentile measure, *N* = 651.

**p < .05. **p < .01. ***p < .001*

REFERENCES

Achtziger, A., & Gollwitzer, P. M. (2010). Motivation and volition in the course of action. In J. Heckhausen (Ed.), *Motivation and action* (pp. 275-299). New York: Cambridge University Press.

Aspen Institute. (2011). *Building teacher evaluation systems: Learning from leading efforts*. Washington, DC: Author.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* White Plains, NY: Longman.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., … Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (Briefing Paper #278). Washington, DC: Economic Policy Institute.

Ballou, D., & Podgursky, M. (1993). Teacher attitudes toward merit pay: Examining conventional wisdom. *Industrial and Labor Relations Review*, *47*(1), 50-61.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1982). Self- efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-147.

Bandura, A. (1993). Perceived self-efficacy in cognitive developing and functioning. *Educational Psychologist, 28*(2), 117-148.

Bandura, A. & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, *88*(1), 87-99.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*(3), 586-598.

Berliner, D. C. (1976). The California Beginning Teacher Evaluation Study: Overview of the ethnographic study. *Journal of Teacher Education, 27*(1), 24-30.

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4): 42-51.

Blair, J. (2002). Gen-Xers apathetic about union label. *Education Week, 21*(20), 1.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011*). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (Working Paper 17699). Cambridge, MA: National Bureau of Economic Research.

Cohen, D. K., & Murnane, R. J. (1985). *The merits of merit pay*. Stanford, CA: Institute for Research on Educational Finance and Governance.

Creswell, J. W., & Clark, V. L. P. (2006). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Darling-Hammond, L., & Berry, B. (1988). *The evolution of teacher policy*. Santa Monica, CA: Rand Corporation.

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis, 23*(1) 57-77.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11,* 227-268.

Dee, T., & Keys, B. J. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management, 23,* 471–488.

Dee, T., & Wyckoff, J. (2013). *Incentives, selection, and teacher performance: Evidence from IMPACT* (Working Paper 19529). Cambridge, MA: National Bureau of Economic Research.

Doherty, K. M., & Jacobs, S. (2013, October). *Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. Washington, DC: National Council on Teacher Quality.

Duckworth, A., Kirby, T., Tsukayama, E., Berstein, H., & Ericsson, K. (2010). Deliberate practice spells success: Why grittier competitors triumph at the National Spelling Bee. *Social Psychological and Personality Science, 2,* 174-181.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087-1101.

Duckworth, A. L., Quinn, P. D., & Seligman, M.E.P. (2009). Positive predictors of teacher effectiveness. *Journal of Positive Psychology, 19,* 540-547.

Eccles, J. A. (2007). Motivational perspective on school achievement: Taking responsibility for learning and teaching. In. R. J. Sternberg and R. F. Subotnik (Eds.), *Optimizing student success in schools with the new three Rs* (pp. 199–202). Charlotte, NC: Information Age.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology* (5th ed., Vol. 3, pp. 1017-1095). New York: Wiley.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2013). *Selecting growth measures for school and teacher evaluations: Should proportionality matter?* (Working Paper 80). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685-706). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2009). Toward a science of exceptional achievement: Attaining superior performance through deliberate practice. *Annals of New York Academy of Science, 1172,* 199-217.

Farkas, S., Johnson, J., & Duffett, A. (2003). *Stand by me: What teachers really think about unions, merit pay and other professional matters*. New York, NY: Public Agenda.

Farkas, S., Johnson, J., & Foleno, T. (2000). *A sense of calling: Who teaches and why*. New York: Public Agenda.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record, 107*(1), 186-213.

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics, 91*, 901-914.

Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal, 44*(3), 594-629.

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. Forthcoming in *Educational Researcher.*

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79*(4), 79-107.

Gist, M. E. (1987). Self-efficacy: Implications for organizational behavior and human resource management. *The Academy of Management Review, 12*(3), 472-485.

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brown Center on Education Policy.

Glazerman, S., & Seifullah, A. (2010). *An evaluation of the Teacher Advancement Program (TAP) in Chicago: Year two impact report*. Princeton, NJ: Mathematica Policy Research, Inc.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*, 3-13.

Goe, L. (2008). *Using value-added models to identify and support highly effective teachers*. Washington, DC: National Comprehensive Center for Teacher Quality.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.

Goldhaber, D. (2009). The politics of teacher pay reform. In M. G. Springer (Ed.), *Performance incentives: Their growing impact on American K-12 education* (pp. 25-42). Washington, DC: Brookings Institution Press.

Goldhaber, D., DeArmond, M. & DeBurgomaster, S. (2007). *Teacher attitudes about compensation reform: Implications for reform implementation* (Working Paper 20). Seattle, WA: Center on Reinventing Public Education.

Goldhaber, D. & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129-145.

Gordon, R., Kane, T. J., & Staiger, D. O. (2006*). Identifying effective teachers using performance on the job* (A Hamilton Project Discussion Paper). Washington, DC: The Brookings Institution.

Hannaway, J., & Rotherham, A. J. *Collective bargaining in education and pay for performance* (Working Paper 2008-11). Nashville, TN: National Center on Performance Incentives.

Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research, 66*(3), 397-409.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (NBER Working Paper No. W11154). Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., & Rivkin, S. G. (2004). How to improve the supply of high quality teachers. *Brookings Papers on Education Policy: 2004*, 7-44.

Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know.* Cambridge, MA: Harvard Education Press.

Heneman III, H. G., Milanowski, A., & Kimball, S. (2007). *Teacher performance pay: Synthesis of plans, research, and guidelines for practice* (CPRE Policy Brief RB-46). Philadelphia, PA: Consortium for Policy Research in Education.

Hershberg, T., & Robertson-Kraft, C. (2009*). A grand bargain for education reform: New rewards and new supports for new accountability*. Cambridge, MA: Harvard Education Press.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher, 41*(2), 56-64.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*(3), 499-534.

Ingersoll, R. M. (2006). *Who control teachers' work? Power and accountability in America's schools.* Cambridge, MA: Harvard University Press.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics, 26*(1), 101-136.

Jacob, B., & Springer, M. G. (2007). *Teacher attitudes on pay for performance: A pilot study* (Working Paper 2007-06). Nashville, TN: National Center on Performance Incentives.

Jennings, J. L., & Corcoran, S. P. (2011). Beyond high-stakes tests: Teacher effects on other educational outcomes. In S. Kelly (Ed.), *Assessing teacher quality: Understanding teacher effects on instruction and achievement* (pp. 77-96). New York, NY: Teachers College Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.). *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York, NY: Guilford Press.

Johnson, S. M. (1984). Merit pay for teachers: A poor prescription for reform. *Harvard Educational Review, 54*(2), 175-186.

Johnson, S. M., Berg, J. H., & Donaldson, M. L. (2005). *Who stays in teaching and why: A review of the literature on teaching retention*. Cambridge, MA: The Project on the Next Generation of Teachers, Harvard Graduate School of Education.

Johnson, S. M., & Papay, J. P. (2009). *Redesigning teacher pay: A system for the next generation of educators*. Washington, DC: Economic Policy Institute.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle, WA: Bill & Melinda Gates Foundation, Measures of Effective Teaching Project.

Kane, T. J., & Staiger, D. O. (2008*). Estimating teacher impacts on student achievement: An experimental evaluation* (NBER Working Paper No. 14607). Cambridge, MA: National Bureau of Economic Research.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation, Measures of Effective Teaching Project.

Kelley, C, Heneman, H. G., III, & Milanowski, A. (2000). *School-based award programs, teacher motivation, and school performance: Findings from a study of three programs*. Philadelphia, PA: Consortium for Policy Research in Education.

Kelly, S. (2011). Understanding teacher effects: Market versus process models of educational improvement. In S. Kelly (Ed.), *Assessing teacher quality: Understanding teacher effects on instruction and achievement* (pp. 7-32). New York, NY: Teachers College Press.

Koedel, C., & Betts, J. R. (2007). *Re-examining the role of teacher quality in the educational production function* (Working Paper 2007-03). Nashville, TN: National Center on Performance Incentives.

Kyriakides, L., & Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19*(2), 183-206.

Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *Future of Children, 17*(1), 87-109.

Lawler, E. E., III (1983). *Pay and organization development*. Reading, MA: Addison-Wesley.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Lockwood, J. R., & McCaffrey, D. F. (2008). *Exploring student-teacher interactions in longitudinal achievement data* (Working Paper 2008-24). Nashville, TN: National Center on Performance Incentives.

Lortie, D. C. (1975). *Schoolteacher*. Chicago, IL: University of Chicago Press.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *MBI: The Maslach Burnout Inventory: MBI*. Palo Alto, CA: Consulting Psychologists Press.

McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.

Meyer, R. H., & Christian, M. S. (2008). *Value-added and other methods for measuring school performance* (Working Paper 2008-17). Nashville, TN: National Center on Performance Incentives.

Milanowski, A. T., Heneman, H., & Kimball, S. (2009). *Review of teaching performance assessments for use in human capital management*. Madison, WI: Strategic Management of Human Capital.

Murnane, R. J. & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review, 56*(1), 1-17.

National Center for Education Statistics. *(2012). Teacher questionnaire: Schools and staffing survey, 2011-2012 school year.* Washington, DC: U.S. Department of Education.

Odden, A. (2008*). New teacher pay structures: The compensation side of the strategic management of human capital*. Madison, WI: Consortium for Policy Research in Education.

Odden, A. & Wallace, M. (2004). Experimenting with teacher compensation. *School Administrator, 61*(9) 24-28.

Odden, A., & Wallace, M. (2007). *How to create world class teacher compensation*. St. Paul, MN: Freeload Press.

Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin Press.

Podgursky, M. J., & Springer, M. G. (2006).*Teacher performance pay: A review* (Working Paper 2006-01). Nashville, TN: National Center on Performance Incentives.

Podgursky, M. J. & Springer, M. G. (2007). Credentials versus performance: Review of the teacher performance pay research. *Peabody Journal of Education, 82*(4), 551-573.

Qazilbash, E. K. (2007, April). *All for one, one for all? Early career teachers' experiences with their teachers' union in an urban district*. Paper presented at meeting of the American Educational Research Association, Chicago, IL.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Robertson-Kraft, C., & Duckworth, A. (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College Record, 116*(3).

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247-252.

Rosenholtz, S. J. (1989). *Teachers' workplace: The social organization of schools*. White Plains, NY: Longman.

Rothstein, J. (2008, April). *Student sorting and bias in value added estimation: Selection on observables and unobservables*. Paper presented at the National Conference on Value-Added Modeling, University of Wisconsin, Madison, WI.

Rothstein, J. (2009). *Teacher quality in educational production: Tracking, decay, and student achievement*. Unpublished manuscript. Industrial Relations Section, Firestone Library, Princeton, NJ.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago, IL: Consortium on Chicago School Research.

Schaufeli, W. B., Leiter, M. P., & Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International, 14*(3), 204-220.

Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains* (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Somech, A. (2010). Participative decision making in schools: A mediating-moderating analytical framework for understanding school and teacher outcomes. *Educational Administration Quarterly, 46*(2), 174-209.

Springer, M.G., Ballou, D., Hamilton, L., Le,V., Lockwood, J.R., McCaffrey, D., … Stecher, B. (2010). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives.

Springer, M. G., Podgursky, M. J., Lewis, J. L., Ehlert, M. W., Ghosh-Dastidar, B., Gronberg, T. J., … Taylor, L. L. (2008). *Texas Educator Excellence Grant (TEEG) program: Year one evaluation report.* Nashville, TN: National Center on Performance Incentives.

Steinberg, M., and Sartain, L. (forthcoming 2014). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project. Forthcoming in *Education Finance and Policy.*

Taylor, E. S., & Tyler, J. H. (2011). *The effect of evaluation on performance: Evidence from longitudinal student achievement of mid-career teachers* (Working Paper 16877). Cambridge, MA: National Bureau of Economic Research.

Teven, J. J. (2007). Teacher temperament: Correlates with teacher caring, burnout, and organizational outcomes. *Communication Education, 56*(3), 382-400.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202–248.

Tyack, D., & Cuban, L. (1997). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

Vroom, V. H. (1964). *Work and motivation*. New York, NY: Wiley.

Walsh, K. (2007). *If wishes were horses: The reality behind teacher quality findings* (Working Paper). Washington, D.C.: National Council on Teacher Quality.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York, NY: The New Teacher Project.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., Davis-Kean, P. (2006). The development of achievement motivation. In N. Eisenberg (Ed.), *Handbook of child psychology* (6th ed., Vol. 3, pp. 933-1002). New York: Wiley.

Wiley, E., Gaertner, M., Spindler, E., & Subert, A. (n.d.). *Denver ProComp evaluation: A mixed-method evaluation of Denver's alternative teacher compensation system*. Boulder, CO: University of Colorado.

Winters, M., Greene, J. P., Ritter, G. & Marsh, R. (2008). *The effect of performance-pay in Little Rock, Arkansas on student achievement* (Working Paper 2008-02). Nashville, TN: National Center on Performance Incentives.