

**Aligning Teacher Improvement Strategies:
A Mixed Method Study of Teaching Reform in Minnesota**

by

Nathaniel L. Schwartz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2012

Doctoral Committee:

Professor David K. Cohen, Chair
Professor Susan M. Dynarski
Professor Brian A. Jacob
Professor Jeffrey Mirel

© Nathaniel L. Schwartz

ACKNOWLEDGMENTS

Many people contributed to this work and I am truly grateful to all of them. I am particularly indebted to my many interview subjects in Minnesota and especially to Kristie Anderson in the Minnesota Department of Education for all the time she spent answering my questions. I am also grateful to Aaron Sojourner, Elton Mykerezi, and Kristine West at the University of Minnesota for their support as I worked to replicate elements of their Q Comp analysis. I would like to offer special thanks to my dissertation adviser, David K. Cohen, who has been a true mentor and friend across the years, and to my wonderful committee members: Susan Dynarski, Brian Jacob, and Jeff Mirel. I benefited from financial support provided by the Rackham Predoctoral Fellowship and data provided through a data award from the Kingsbury Center of the Northwest Evaluation Association. Finally, Seneca Rosenberg served as chief discussant, critic, and advocate through every stage of this project and I cannot thank her enough.

TABLE OF CONTENTS

Acknowledgments.....	ii
List of Figures.....	iv
List of Tables.....	v
Abstract.....	vi
Chapter One Introduction.....	1
Chapter Two What can we expect from teacher improvement systems? A quantitative evaluation of Q Comp in Minnesota.....	10
Chapter Three Aligning teacher improvement strategies at the state level: The design and implementation of Q Comp in Minnesota.....	72
Chapter Four Defining instructional expertise: Patterns of district variation in Q Comp implementation.....	117
Chapter Five Conclusion.....	177

LIST OF FIGURES

Figure 2.1: Number of Q Comp Districts by Year.....	56
Figure 2.2: MCA and NWEA Test Trends by Q Comp Adoption Cohort.....	57
Figure 2.3: Subject Focus Distribution.....	58
Figure 2.4: Test Focus Distribution.....	58
Figure 2.5: Program Type Distribution.....	58
Figure 3.1: Teacher Improvement System Schematic.....	76
Figure 3.2: Traditional School Districts in Q Comp by Year.....	88

LIST OF TABLES

Table 2.1: Baseline Descriptives.....	59
Table 2.2: Comparison of Full and Matched Sample.....	60
Table 2.3: Q Comp Effects on Teacher Salary.....	61
Table 2.4: Q Comp Effects on Professional Development.....	62
Table 2.5: Q Comp Effects on Teaching Climate.....	63
Table 2.6: Q Comp Effects on Student Achievement.....	64
Table 2.7: Q Comp Effects across Achievement Distribution.....	65
Table 2.8: Q Comp Effects across Student Subgroups.....	66
Table 2.9: Q Comp Effects by Program's Subject Focus.....	68
Table 2.10: Q Comp Effects by Program's Test Focus.....	69
Table 2.11: Summary Statistics by Program Type.....	70
Table 2.12: Q Comp Effects by Program Type.....	71
Table 4.1: District Descriptives Before Q Comp Adoption.....	128

ABSTRACT

ALIGNING TEACHER IMPROVEMENT STRATEGIES: A MIXED METHOD STUDY OF TEACHING REFORM IN MINNESOTA

Despite consensus that teachers vary considerably in their effectiveness, researchers have had little success identifying strategies to increase teacher quality at scale. Policymakers often call for isolated reforms such as teacher performance bonuses, notwithstanding the poor track record of previous efforts. This dissertation investigates a new generation of human capital programs that combine incentive reform with teacher development. These efforts enjoy support from research within a variety of fields. If teachers do not have the capabilities to improve, revised incentives are unlikely to cause meaningful change, just as increased opportunities to learn will not translate into better instruction if teachers lack incentives to apply concerted effort to this task. However, we know little about how to translate such research into designs for teacher and student learning.

I draw on evidence from a program in Minnesota known as Quality Compensation for Teachers or “Q Comp.” The program, which began in 2005, includes many features that can be found in several more recent state-level reforms, including requirements for teacher evaluation, pay reform, and teacher-led professional development.

The first essay evaluates the impacts of Q Comp on school practices and student achievement, drawing on several sources of quantitative data. While I find that Q Comp caused changes to teacher-reported school practices, I find little evidence that the program produced meaningful improvement in student learning. The second essay considers program

implementation at the state level, describing the interactions between the legislature, the union, and the Minnesota Department of Education, which created a program that included relatively effective structural oversight with weak instructional guidance and training. The final essay focuses on district implementation, illustrating how three districts' design choices around teacher support and development produced programs that differed considerably in their ultimate significance for teachers even as they maintained a rough structural similarity.

As a group, these essays explore a series of concerns associated with a new generation of state reform programs. They raise doubts about the likelihood that these programs will lead to large-scale instructional improvement, and they illustrate the difficulty of setting up large-scale human capital systems at the state level.

CHAPTER ONE

INTRODUCTION

Educational policy itself continues to grow: reformers in and out of government are devising more demanding standards, designing more challenging assessments, writing more thoughtful instructional frameworks, and developing more intelligent curricula. Each of these endeavors increases the things that teachers would have to learn if they were to succeed, but few are well enough designed to promote such learning. To make such policies more educational for educators, the processes in which policy is made and enacted would have to be opened up so that they created many more opportunities to learn. And those opportunities would have to be designed so that they embodied the sorts of teaching and learning that reformers wish to promote for classrooms.

—Cohen & Barnes, 1993

The last several years have been a busy time for reformers concerned with teacher quality. Policymakers at both ends of the political spectrum have made teachers the central focus of their interventions, proposing plans to change every part of the teacher work cycle, from training on through hiring, paying, developing, and firing.

Within the set of proposed and already passed reforms, one can identify two broad trends. First, teacher policy seems on track to become increasingly punitive toward teachers who fail to perform according to specified criteria. Teachers will undergo more measurement and evaluation, allowing supervisors to sort the pool into categories of effectiveness. And weakened tenure and pay protections will make it easier to force teachers who are deemed ineffective out of the profession and to reward those teachers who perform at higher levels. By raising standards around performance and pushing teachers out who do not meet those standards, such policies could fulfill an important sorting function, leading to a better qualified workforce (Hanushek,

2009; Staiger & Rockoff, 2010). But these policies also seem to suffer from the issue identified by Cohen and Barnes in the quotation that begins this chapter. They raise expectations for teachers without necessarily providing the kind of learning opportunities that might help incumbents learn to meet these new expectations. If the aim is to raise teacher quality in part by improving the performance of teachers already in the classroom, incentive-based policies alone are unlikely to be sufficient.

Second, in a sometimes uncomfortable juxtaposition with the first trend, support has grown for a host of policies aimed at providing teachers with new kinds of instructional feedback and professional development to spur on-the-job improvement. Here, we see a renewed interest in strategies ranging from peer observation and coaching to teacher learning communities to standardized professional development classes. Such initiatives are popular in themselves but have also gained steam as a consequence of the move in many states to implement Common Core standards. A cynical way of looking at this second set of teacher learning reforms might be to see them as political maneuvers designed to make already described punitive measures more acceptable to teachers and to the general public. At the same time, ideas about incorporating better opportunities for teacher learning into schools and districts have a long history of their own, and they seem increasingly necessary as a counterpoint to the push for teacher accountability (Darling-Hammond & McLaughlin, 1995; Hargreaves & Fullan, 2012). Recent acclaim for the Finnish system of education after the nation received top marks on the Programme for International Assessment (PISA) test has added support in this direction, since Finland's educational system appears to place a premium on teacher professional development over accountability (Sahlberg, 2011).

The two reform trends come together in the recent set of state Race to the Top (RTTT) proposals. The RTTT asked states to submit to the federal government plans for teacher-focused policy reforms that revise teacher evaluation, compensation, and retention policies to “encourage and reward effectiveness” and that expand “effective support to teachers and principals,” making available \$4.35 billion in competitive funding to the winners of the grant (White House Press Office, 2009). States’ proposals in turn have incorporated a host of teacher effectiveness strategies, including widespread teacher evaluation, reforms to link pay, promotion, and dismissal to teacher performance, and a series of feedback and professional development initiatives. Much of the political controversy around these plans has tended to be focused on the means that states will use to sort teachers into different groups based on effectiveness and the repercussions for teachers who miss the mark. Perhaps as a result of these controversies, many states tempered to some degree the emphasis on pushing teachers out of the classroom and are increasingly focused on changing the system of incentives, feedback, and rewards for the larger pool of teachers who exist between the extremes of least and most effective (Jerald, 2012).

In this dissertation, I define the term teacher improvement system to capture the new generation of programs aimed at these types of workforce improvement. I propose that such programs are designed to simultaneously address two related concerns. On the one hand, they attempt to shift the thrust of teachers’ professional incentives through a combination of career ladder and pay opportunities to increase the focus on student learning. On the other, they hope to alter teacher learning opportunities in order to make it easier for teachers to meet their professional goals.

By connecting these elements within a single policy, these programs might contribute to a more coherent policy landscape, addressing the frequently raised concern that the teacher

development system in its current incarnation generates a hodge-podge of divergent incentives and supports and thus leaves teachers without many of the resources necessary to build a stronger professional practice (Odden, 2008; Bryk et al., 2009; Cohen & Moffit, 2009, Hargreaves & Fullan, 2012). Moreover, such programs, which combine raised teacher expectations with various mechanisms for teacher training, seem like they might address the need raised by Cohen and Barnes in the opening quotation for programs that make provision for teacher learning alongside other policy changes.

However there has been little research into the design or operation of these programs. Driven in part by the promise of federal dollars and in part by the favorable political atmosphere, state governors, legislatures, and departments of education are racing to rapidly roll-out fully realized systems with little testing or experimentation. Notwithstanding the broad arguments sketched above in favor of such policies, these developments are concerning. They are concerning because governments are spending time and resources that might potentially go toward more productive programs if such initiatives are found to be ineffective, and they are concerning because many of the design issues going untested are likely to be crucial to the ultimate outcome of such systems.

As a rule, state agencies have proven quite ineffective at managing large-scale instructional reform in ways that create substantial changes to instructional practice or student learning. The state efforts that have arguably seen the greatest success have tended to be test-based accountability efforts where the state has not played a major role in instructional support (Dee & Jacob, 2011). In contrast, state efforts to deliver standardized ideas about instruction have often foundered due to lack of capacity and expertise (Spillane, 1999; Cohen & Spillane, 1993; Smith & O'Day, 1990). It is not clear where teacher improvement systems will fall on this

continuum or how states and districts might balance the responsibility for these programs. The ways these tensions unfold in practice will play an important role in determining the programs' viability.

Given these concerns, what are the prospects for such reforms and what might they take to succeed? I address these questions in the following dissertation through a series of articles on a program in Minnesota called Q Comp. Following its launch by the Minnesota Department of Education in 2005, the program has been picked up by over 80 districts, affecting around one-third of the state's public school students. What makes this program particularly interesting is how closely it resembles many of the state reforms being designed today. Like the RTTT designs, Q Comp is a state-directed program that asks districts to implement within state guidelines plans that include teacher evaluation, performance pay, teacher career ladders, and professional development led by teacher-leaders. Unlike the more recent initiatives, Minnesota's program has been offered as a choice to districts rather than a mandate, but the shape of the program and the breadth of adoption still mean that the program offers the best available evidence on how such teacher improvement systems might unfold.

My first essay provides a quantitative evaluation of Q Comp's effects on district practice and student achievement. Using a comparative interrupted time series model to identify the program's causal impacts, I draw on data from the Schools and Staffing Survey, the Minnesota Department of Education, and the Northwest Evaluation Association. I find that Q Comp changed several aspects of district practice, including teachers' pay, their professional development, and their perceptions of the teaching climate in their schools. At the same time, I find little to no overall impact on student learning outcomes, although there is evidence that the program shifted the achievement distribution, raising scores for the lowest achievers while

lowering scores for those who entered with higher test scores. I further investigate the relationship between district effects and district program designs and find weak evidence that certain program types might be associated with greater student gains.

The second essay begins to explain Q Comp's successes and failures through an analysis of program operations. I begin by situating teacher improvement systems within the context of large-scale instructional reform and speculating about what it might take to make these programs meaningful sources of improvement. I then use these ideas to explore Q Comp as it developed at the state level. Through interviews with legislators, state officials, and state union representatives, I work to explain the influences that shaped the program and the ways that politics and capabilities defined the program in legislation and practice. In particular, the study looks to understand the program's transition from the legislature to the state department, and the strategies and coping mechanisms that the state department adopted for implementing a program unlike any others previously in its charge. I conclude that these interactions produced a program that offered districts strong guidance around the structural design of the program but little direction about the types of content the program should deliver. When districts did receive guidance around program content, it tended to come from the union rather than the department of education. I argue that the ambiguity around program content created challenging conditions for the development of a robust, statewide improvement system.

In the third essay, I move from the state to the district level, using a comparative case study approach to investigate the ways that state patterns of oversight and guidance affected district program design choices. I explore sources of variation and similarity and their implications for program operations. There is a considerable literature on the role that districts play in shaping state and federal reforms, but this literature has not fully explored the role that

districts can play in defining what is meant by instructional learning within programs like Q Comp. I argue that the Q Comp designs in each of my three case study districts shared a host of structural features, however the districts took very different approaches to integrate instructional expertise into their programs. This in turn meant that the programs offered different sorts of teacher supports and had different prospects for building the kind of knowledge base that one would hope to find within a teacher improvement system. I argue that discussions of district capacity to implement such programs must take into account the types of internal and external connections that districts have to new instructional ideas, as well as the guidance they are receiving from state agencies.

As a group, these essays explore a series of concerns associated with the new generation of state reform programs. They raise doubts about the likelihood that these programs will lead to large-scale instructional improvement, and they illustrate the difficulty of setting up large-scale human capital systems that include the kinds of learning that Cohen and Barnes described in the quotation that opened this chapter.

REFERENCES

- Bryk, A.S., Sebring, P.B., Allensworth, E., Luppescu, S., and Easton, J.Q. (2009). *Organizing Schools for Improvement: Lessons from Chicago*. Chicago: The University of Chicago Press.
- Cohen, D. K., & Barnes, C. A. (1993). "Conclusion: A new pedagogy for policy?" In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.
- Cohen, D. K. and Moffitt, S. L. (2009). *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Cambridge, MA: Harvard University Press.
- Darling-Hammond, L. & McLaughlin, M. (1995). "Policies that support professional development in an era of reform." *Phi Delta Kappan*. 76(8): 597-604.
- Dee, T. S. & Jacob, B. A. (2011). "The impact of No Child Left Behind on student achievement." *Journal of Policy Analysis and Management*. 30(3): 418-446.
- Hargreaves, A. and Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York: Teachers College Press.
- Hanushek, E.A. (2009). "Teacher Deselection." In *Creating a New Teaching Profession*, edited by Dan Goldhaber and Jane Hannaway (165–80). Washington, DC: Urban Institute Press.
- Jerald, C. (2012). "Movin' it and improvin' it: Using both education strategies to increase teaching effectiveness." Center for American Progress. Accessed online at http://www.americanprogress.org/issues/2012/01/movin_it_improvin_it.html on June 29, 2012.
- Odden, A. (2008). "New teacher pay structures: The compensation side of the strategic management of human capital." Consortium for Policy Research in Education.
- Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?* New York: Teachers College Press.
- Spillane, J. P. (1999). "State and local government relations in the era of standards-based reform: Standards, state policy instruments, and local instructional policy making." *Educational Policy*. 13(4): 546-572.
- Smith, M. S., & O'Day, J. (1990). "Systemic school reform." In Fuhrman & B. Malen (Eds.). *The politics of curriculum and testing: The 1990 yearbook of the Politics of Education Association*. London: Taylor & Francis.
- Staiger, D. O. and Rockoff, J. E. (2010). "Searching for effective teachers with imperfect information." *Journal of Economic Perspectives*. 24(3): 97-118.

White House Press Office. (2009). *Race to the Top Fact Sheet*. Accessed online at <http://www.whitehouse.gov/the-press-office/fact-sheet-race-top> on July 5, 2012.

CHAPTER TWO

WHAT CAN WE EXPECT FROM STATE TEACHER IMPROVEMENT SYSTEMS? A QUANTITATIVE EVALUATION OF Q COMP IN MINNESOTA

I. INTRODUCTION

Over the next few years, most states across the nation will undertake wide-ranging teacher quality reforms (NCTQ, 2011). These reforms, encouraged by the federal Race to the Top (RTTT) grant competition, involve state-level initiatives to align district strategies around teacher evaluation, pay, and professional development. The teacher improvement systems now in development represent a new frontier of state educational policy, transforming the state's role with respect to teacher human capital policy. There are several studies that look at the impact of the individual components that will make up these new systems, but few studies consider the impact of the broader state-level programs. Moreover, the evidence about the different components is mixed, suggesting that much about the integrated programs will depend upon their particular design and implementation processes.

This paper examines the impacts of a close predecessor to these programs. The Q Comp program, which launched in Minnesota in 2005, is strikingly similar to the initiatives now in development in many states because of the RTTT. Under Q Comp, Minnesota funds and to some extent trains districts to adopt a series of interconnected human capital reforms. The requirements for district entrance into the program include a teacher evaluation system, teacher leadership opportunities, and performance-based pay raises. As the first and only fully developed

example of a state-led teacher improvement system currently in operation, this program offers an important source of evidence about the new wave of state reform.

My empirical analysis is based on several sources of data, including teacher responses from the Schools and Staffing Survey, school-level state testing data provided by the state of Minnesota, and, for a subsample of schools, student-level data for an alternate benchmark assessment provided by the Northwest Evaluation Association. The analysis builds upon the only other causal study of Q Comp's effects, a study recently completed by Sojourner, Mykerezi, and West (2012). The authors take a similar approach to the data, using difference-in-difference and interrupted time series analysis, to try to isolate the effects of Q Comp on adopting districts. They find small overall program effect on student achievement, along with inconclusive evidence of heterogeneous effects by district program characteristics.

While the Sojourner et al. analysis provides an important source of data about Q Comp, the authors consider Q Comp primarily as a performance pay program and thus confine their focus to only certain aspects of the state initiative. For instance, their consideration of program variation across districts is limited to differences in bonus payments offered to teachers, making their findings somewhat difficult to interpret.

My contributions fall into three different areas. First, I offer new evidence about the extent of program implementation at the district level using teacher-level data from the Schools and Staffing Survey. This evidence demonstrates that districts that joined Q Comp did make small changes to teacher salary schedules as the program demanded. But it also shows that districts changed in other ways associated with the broader TIS program, and it demonstrates the need for an analysis of Q Comp that looks beyond the merit pay aspects of the program.

Second, I provide further verification of Sojourner et al's conclusion that the Q Comp program produced small if any overall gains in student achievement across adopter districts using a slightly different model that adds controls for pretrends in district achievement. However, I do find intriguing though still preliminary evidence that the program caused pronounced shifts in the student achievement distribution, raising scores by more than 0.2 standard deviations for students who previous to the program fell into the lowest achievement quartile even as the program caused a similarly sized drop in scores for students who began in the highest quartile. Interestingly, I do not find similar differences in program effect when I cut the data by other observable student characteristics such as race, gender, or SES.

Third, I attempt to understand differences in effect across districts and schools. Where Sojourner et al. looked solely at differences in performance pay offerings across districts and found little evidence of meaningful variation in effect, I focus on two other reasons for effect heterogeneity. Since schools were required by Q Comp to set yearly student achievement goals in either math or reading using either state or Northwest Evaluation Association assessments, I test the hypothesis that Q Comp achievement gains were concentrated in particular focus areas or test results during different years and still find little evidence of a large overall achievement effect. I also look at differences in district-level supports for teacher learning within the program, a choice driven by previous qualitative study (Schwartz, 2012). In particular, I divide districts by the emphasis of the instructional development within their Q Comp programs. I find some evidence of a positive and meaningful impact on math achievement in districts that focused on bringing in external instructional expertise as a part of their Q Comp design, however the effect appears to be driven by a small number of districts and does not show up equally on reading achievement.

These findings offer crucial insight into the possibilities and challenges for statewide teacher improvement systems. The large gains due to Q Comp for students who began at the bottom of the achievement distribution speak to the potential of such programs. At the same time, the losses for higher achieving students raises serious questions about the route by which these changes occurred. The pattern of results suggests that Q Comp did not produce the overall instructional improvements that were expected – the rising tide that would help all students – and instead caused a redistribution of teacher effort, technique, or attention in ways that helped some students and hurt others. This absence of meaningful overall effects speaks to the challenges that states will face as they move to launch programs that closely resemble Minnesota’s Q Comp.

While my investigation into effect heterogeneity by district program characteristic suggests some factors to consider as states make design choices about the new programs, the evidence here is still quite inconclusive about what would make these programs more successful at raising student scores across the board. These findings should be troubling to state policy-makers who have devoted significant state and federal resources to the design and implementation of the new systems.

The remainder of this paper is structured as follows. Section 2 discusses previous literature and briefly sketches the theory behind teacher improvement systems. Section 3 provides details on Q Comp as a state program. Section 4 outlines the data and methods used in this study, and Section 5 describes the results. The final section concludes.

II. PREVIOUS LITERATURE

With teacher effectiveness at the forefront of most educational reform efforts, teacher improvement systems have become an increasingly popular policy response to the problem of

variability in teacher quality. Several high profile districts now engage programs that draw together a series of potentially complementary human capital strategies. These include Washington D.C. and its IMPACT program, Denver Colorado, which runs ProComp, as well as the more than 50 districts that implement the Teacher Advancement Program. Such programs have more recently gained favor at the state level, encouraged by both state policy debates and federal programs like the Race to the Top. Proponents of these state-level policies suggest the reforms might both shift the make-up of the teacher workforce in favor of more effective teachers and generate improvement among those teachers already on the job. However, since these new systems include multiple interconnected components, assessing their potential based on previous research is difficult.

Media attention to many of these new policy plans often focuses almost exclusively on their attempts to link teacher pay or dismissal to student outcomes (Sawchuk, 2009). In contrast with the evidence on teacher improvement system reform packages, recent evidence on pay for test scores is remarkably consistent. A series of highly rigorous randomized experiments conducted over the past few years suggest that pay-for-performance systems when divorced from other human capital reforms are unlikely to yield substantial benefits in student achievement or teacher practice. These studies include the Tennessee POINT experiment that tested the impact of individual-level bonuses, an experiment in Red Rock, Texas looking at team-based payments, and a study of school-wide bonuses in New York City. In each case, the bonuses offered for raising student scores were substantial and the responses in terms of student achievement and teacher behaviors was minimal to non-existent (Springer et al., 2010; Springer et al., 2011; Fryer 2011; Marsh et al., 2011). While the studies did not test every possible policy iteration associated

with pay-for-performance, their combined conclusions provide considerable reason to question teacher outcome-based bonuses as a central means of improvement.

Yet what such bonus-pay studies say about broader human capital reforms remains unclear. Teacher improvement initiatives that make pay differentiation only one piece of the system offer several compelling reasons for preferring them to simple merit pay.

First, there is some indication that bonus pay alone might fail because teachers do not know how to best adapt their instruction to produce better outcomes. If this is the case, shifting incentives might have little impact even if they produce increased effort. Research on incentive pay outside the field of education offers myriad examples of this phenomenon. Camerer and Hogarth (1999), for instance, review studies on lab-based incentive experiments and conclude that incentive effects depend largely on the alignment between the subject's skills and knowledge and the demands of the task at hand. In operations management, the same relationship is characterized by the "motivation-opportunity-ability" framework that describes the ways that an individual's goals are mediated by context so that outcomes depend on the interplay between the individual's aims, his environment, and the skills and knowledge that he has available (Siemson et al., 2008). Applied to teaching, these theories offer one possible explanation for the failure of rewards-based strategies to produce measurable effects, namely that many teachers in the current environment are unable to determine how to productively shift instruction to raise achievement outcomes. Several qualitative researchers have explored the lack of coherent instructional feedback and learning opportunities and the confusion that this can cause among teachers who are attempting to change their instructional practice (Spillane & Jennings, 1997; Coburn, 2001).

Another strand of evidence in favor of the current set of teacher reform initiatives comes from research into systemic reform in education. The thrust of this literature is that teacher improvement efforts tend to be hampered by a highly fragmented system of governance and resources, where there is little alignment between classroom feedback, teacher advancement, professional development, or school goals (e.g. Miles et al., 2002; Newmann et al., 2001). Cohen and Spillane (1992), for instance, in a study of the relations between governance and instruction in the U.S. note that the responsibility for teachers' instructional guidance is divided among a variety of local sources and agencies, creating an outpouring of guidance that is often "unrelated, divergent, or contradictory" (p. 22). In the context of an instructional environment where there are few shared resources or incentives and little direct feedback, large-scale teacher improvement becomes highly unlikely (Cohen & Moffitt, 2009). There is growing evidence that several of the interventions with evidence of a consistently positive student achievement effect across multiple schools and classrooms including some comprehensive-school reforms like Success for All and America's Choice and charter networks like KIPP and Achievement First are those that have developed coherent and protected guidance systems for their teachers in the midst of the broader context (Borman et al., 2007; Correnti & Rowan, 2007; Angrist et al., 2010; Rosenberg, 2012). Packages of human capital reforms such as those driven by the Race to the Top can be viewed as a similar attempt to increase coherence by standardizing teachers' goals and instructional feedback across classrooms, schools, districts and states.

Finally, the set of packaged human capital reforms associated with Race to the Top supplement pay reform with a series of reform initiatives including teacher observation and evaluation, targeted professional development, and data-based feedback that are promising in their own right and that might additionally benefit from complementarities across elements.

While the full literature in each of these areas is beyond the scope of this review, several recent high-quality studies demonstrate the potential of these types of reforms. Taylor and Tyler (2011), for instance, find that classroom-based observation and evaluation can serve as a useful professional development tool, raising the performance of mid-career teachers by around 10 percent of a standard deviation both during the evaluation period and in subsequent years following the evaluation. Glazerman et al. (2010) in a randomized study of comprehensive teacher induction measures that included several types of support and coaching, find that two-year induction programs generate positive and significant student test score improvements. And a recently released Gates Foundation study (2012) determines that several currently popular frameworks used for teacher observation and evaluation can accurately identify teachers who are most effective in raising student outcomes and those in need of remediation. At the least, these studies suggest that there might be benefit to adopting teacher improvement reforms for their individual elements as well as for the greater package.

However, from the standpoint of program design, these reform packages are quite complex. Even with some evidence that the individual elements in the packages can make a difference to teacher behavior and student outcomes, there remains considerable debate about how each of the elements should be designed and implemented. Creating a system that allows both successful implementation of each part along with coherence across reforms as well as integration with the other already existing teaching improvement efforts in the district is a difficult endeavor. Moreover, since some of the individual elements within the new package reforms are popular in their own right, it is unclear whether a state-mandated adoption of the entire package would be more effective than the district-by-district and school-by-school adoption of individual elements that is currently taking place.

Current knowledge of teacher improvement systems is limited. Some of the best available evidence comes from studies on the Teacher Advancement Program, which works with schools and districts to implement a system of teacher evaluation and pay rewards. A randomized study of the program in Chicago that included only eight treatment schools and two years of data found no evidence of improvement in student test scores (Glazerman & Seifullah, 2010; Glazerman et al., 2009). While these results are informative, their small sample size and lack of long-term data suggests the need for caution in generalizing the results. Two other studies offer some further data about TAP's effects. Springer et al. (2008) use longitudinal panel data from two unidentified states to look at the effects of TAP on student achievement. While this study is not confined to a small geographic area, the sample still includes only 13 TAP schools at its largest point. The authors find that the program significantly increased student test scores in the elementary grades but that the effects disappeared at the middle and higher grades. Solomon et al. (2007), in a TAP-funded study, match a wide sample of TAP schools and teachers with a convenience sample, raising the possibility of important unobservable differences between treatment and comparison schools. The authors found that students in TAP classrooms looked significantly more successful than students in control classrooms in a wide range of grades, states, and outcomes.

It is difficult to know how these mixed TAP results might generalize to the state reforms currently in progress as a result of Race to the Top. TAP is an externally designed program implemented at the district level. In contrast, the state departments of education that are now in the process of implementing similar reforms face a different set of constraints and opportunities. Since most of the state-level reforms are still in the process of development, however, there has been relatively little research into these programs.

In the next section, I describe Minnesota's Q Comp program in detail. Unlike some of newer systems, the program allowed districts to choose whether to opt in, creating a program with certain important differences from some of the Race to the Top initiatives. But I will argue that for the most part, Q Comp closely resembles more recent TIS designs and thus provides an important source of evidence as the only fully developed, state-led TIS currently in operation. As noted in the introduction, Sojourner et al. offer the only other rigorous study of program outcomes. The authors find some evidence of a small overall program effect on achievement (~ 0.03 student standard deviations) with little evidence that differences in district allotment of pay bonuses are associated with differential program effects. I replicate several of the Sojourner models in the following paper, although Sojourner et al. benefit from additional access to student-level state testing data.

III. PROGRAM DESCRIPTION

The Minnesota statute that created Q Comp in 2005 gives the Minnesota Department of Education (MDE) the power to grant state funding to traditional and charter districts for program designs that include each of the following elements:

- Salary schedule reform connecting a portion of teacher pay increases to performance. Acceptable teacher performance benchmarks include schoolwide achievement gains, measures of student achievement at the teacher level, and/or teacher evaluation results.
- Career advancement opportunities for teachers that allow teachers to retain primary roles in student instruction.
- Site-based professional development, led by master or mentor teachers, and taking place during the school day.

- Site and district-level school performance goals (in math or reading) that serve as the basis for schoolwide performance benchmarks and site-based professional development.
- Joint sign-off from local superintendent and local union president (or teacher representative in case of non-unionized charters)

The design of Minnesota's program resembles the current wave of state policy changes in several important ways.

First, teacher pay is only a single element of the greater program of reform. Indeed, only one paragraph of the relevant statute actually concerns teacher pay; the rest of the several pages of legislation specify the pieces relates to professional development, evaluation, and teacher leadership that must be included in district plans. Second, districts have wide flexibility to define the details of the program, especially around many of the teacher learning components. Across the set of program designs, salary schedules tend to look relatively similar, but districts exercised considerable autonomy in setting up their teacher career ladders and professional development components. Third, the state's investment in the program has been fairly substantial overall – over \$200 million – but is small on a per-pupil level. The grant to adopter districts of \$190 per student in state aid and up to \$70 per student in partially equalized levy represents only around one percent of the average per-pupil spending across the state. Fourth, the capacity of the state department of education to support district Q Comp designs has been severely limited by state funding crises and a lack of designated staff in the central office who worked solely with the Q Comp program (Schwartz, 2012).

The combination of high district autonomy and weak state support raises questions about the degree to which Q Comp represents a defined program of intervention. Because similar program designs are now being rolled out across the US with a huge outpouring of time and

resources devoted to the task, one might argue that program evaluation is necessary even though the outline of the intervention is somewhat blurry. However, my analysis is not confined to the program's overall impacts on achievement. In the initial results section, I investigate the evidence that districts actually changed practices due to the adoption of Q Comp using the Schools and Staffing Survey as evidence. Then, in a later part of the section, I attempt to capture the effects of district autonomy by looking at interactions between the adoption of Q Comp and the particulars of district's individual program designs.

IV. METHODS AND DATA

A. Empirical Methodology

My analysis strategy across outcomes is similar to Sojourner et al. (2012), in that I use a treatment-comparison research design where I compare districts that adopted Q Comp in a particular year to districts that adopted the program later in time or to districts that never adopted at all. By changing the comparison group, I am able to test the sensitivity of results to my choice of controls (Dynarski, 2008).

The simplest version of this specification is a variant on a difference-in-difference (DD) analysis to take into account the different program adoption years in different districts. Here, the first difference is in average outcomes before and after adoption for the treatment group and the second difference looks at the change in outcomes over the same time period for the control group.

$$(1) \quad y_{gst} = \beta_1 Q_{gdt} + \gamma X_{gst} + \alpha_{gt} + \alpha_{gd} + \varepsilon_{gdt}$$

In this specification, y denotes the outcome, shown here at the grade-school-year level, although some of my outcomes are available at the student or teacher level. Q serves as a dummy treatment indicator that turns on for particular districts in years after they have implemented Q-Comp. X is a vector of district demographic variables that control for compositional changes in a district's student body. The terms α_{gt} and α_{gd} in the equation represent grade-district and grade-year fixed effects that ensure that the treatment effect is estimated separately for each district, grade, and time period. Note however that the level of fixed effects I include (student, school-grade, etc.) depend on the particular dataset that I use, as described in the next sections. When student level data is available, I include student and year fixed effects instead. ε represents a general error term for all other unexplained variance. For this analysis, and all following, standard errors are robust and clustered at the level of treatment (e.g. the district level) to allow for correlation between observations within districts.

The major threat to causal identification in this type of analysis is that outcomes change in ways that appear to be attributable to a program effect but actually are due to some other change that is correlated in timing with the program of interest. For instance, differential shifts between treatment and control districts in student body composition might change test scores but be either unrelated to or even the cause of Q Comp adoption. One advantage for this study is that Q Comp has been implemented in different districts in different years (Figure 2.1). This program roll-out scheme makes it less likely that my results will be significantly biased by a single external event. However, as shown above, most of my specifications include controls for all available student body characteristics, including such things as student race, LEP and free/reduced-price lunch classification.

Since the SASS data that I use to investigate program implementation only included two survey waves, one before the start of Q Comp and one after a number of districts had adopted the program, the simple DD analysis shown above is the only one possible. However, student testing data is available yearly, allowing model specifications that take into account time trends before and after program adoption.

Sojourner et al. (2012) offer a specification shown below that includes a dummy control variable for district-years that take place two or more years before adoption.

$$(2) \quad y_{gst} = \beta_1 Q_{gdt} + \beta_2 PREYEAR_{gdt} + \gamma X_{gst} + \alpha_{gt} + \alpha_{gd} + \varepsilon_{gdt}$$

The addition of the dummy variable *PREYEAR* alters the comparison so that average outcomes after program adoption are now compared to average outcomes in the year directly before adoption. This specification provides a basic check on the assumptions of the model. A significant coefficient on the *PREYEAR* dummy variable implies that district outcomes changed significantly in the year before adoption as compared with previous years, calling into question the assumption that program adoption is as good as random when conditioned on the included covariates.

Importantly, however, Specification 2 does not take into account the direction that outcomes were trending before program adoption. This matters because higher average post-adoption outcomes in a particular district might be explained by a positive trend in the pre-data, just as lower outcomes might be due to sinking test scores overall rather than Q Comp adoption. In Specification 3, I control for district pretrends. To do this, I estimate regression coefficients for each district on the relationship between the test outcome and the year using only outcomes

and years from before Q Comp adoption. I then interact these coefficients with all study years and add these interactions as covariates in the model (see Dynarski, 2004):

$$(3) \quad y_{gst} = \beta_1 Q_{gdt} + \beta_2 PRETREND_d + \gamma X_{gst} + \alpha_{gt} + \alpha_{gd} + \varepsilon_{gdt}$$

As depicted by the subscripts on pretrend, trends are calculated individually for each district in the years before the adoption of Q Comp. I then add controls for the district's estimated coefficient*year across the panel of all years. To make the initial regressions possible, I limit the sample that I use to only those districts that include at least two years of predata.¹

B. Sample and Summary Statistics

Between 2006 and 2010, 49 traditional districts joined Q Comp bringing a total of around 260,000 students into the program.² While 49 districts represents only around 15 percent of the 364 total traditional districts in Minnesota, many of Minnesota's districts are quite small and the 49 that joined Q Comp encompass around one-third of the state's student body. During the study time period, one district voluntarily withdrew from the program and one district was expelled from the program by the state for non-compliance. Charter districts were also allowed to join the program, and around 25 single-school charters did join over the study period, however I leave charters out of my analysis since (a) the treatment in these schools is even less clear than for

¹ Note that none of these models distinguish between earlier and later years of implementation. It is possible that a Q Comp effect would only be detectable after several years of program operation. In unpublished results, I ran the same models with dummies for implementation year. These models significantly decrease power, and standard errors become quite large. However, there was no indication of significant differences between early and later adoption years.

² While most Q Comp adoption took place at the district level, a handful of schools in Minneapolis and Roseville districts joined the program even while the rest of their districts remained outside Q Comp. The analyses I present here allow for Q Comp coding at the school-level. Removing these schools/districts from the analysis does not significantly change results.

traditional districts since charters face fewer initial restrictions previously on teacher pay policies; (b) the demands of implementing Q Comp across an entire district, which have been the focus in the other chapters of this dissertation, are considerably different than the demands to implement in a single charter school, and; (c) no Q Comp charter schools use the NWEA test that I use for one of my outcomes, constraining the type of data available. Figure 2.1 shows cumulative number of traditional districts into Q Comp by year.³ As is evident in the figure, the largest entrance waves into the program occurred during the 2007-2008 and 2008-2009 school years.

My primary data source for analysis comes from enrollment records provided for all schools in Minnesota by the Minnesota Department of Education. These files provide school-grade level information on all students in the Minnesota public system, including most basic demographics. As noted, I restrict the data to traditional districts and to regular schools within these districts, as defined by the Department of Education. I match these with school-grade test information from the state as well as with student-level test information provided by the Northwest Evaluation Association (see below). I additionally collected district-level funding information from the Common Core of Data and teacher-level survey responses from the Schools and Staffing Survey.

Table 2.1 presents summary statistics from the 2004-2005 school year (the year before Q Comp began). The first column shows district averages for districts that would join Q Comp over the next five years and the second column shows averages for those that stayed out of the program. As the table demonstrates, there are significant differences between adopter and non-adopter districts. Districts that joined Q Comp tended to be far larger, enrolling several thousand

³ As shown in Figure 1, the majority both of districts and students to join Q Comp were in non-charter settings. My full analysis includes both types of schools, however analyses that compare state and NWEA tests look only at traditional districts, since I have no NWEA charter school data.

more students than non-Q Comp districts. Q Comp districts also tended to be better off. These districts had a lower percentage of free- and reduced-meal students, a higher percentage of teachers with masters degrees, and they correspondingly paid higher teacher salaries. Similarly, Q Comp districts averaged higher scores on state tests than non-Q Comp districts by around one-half of a standard deviation.

The extent of the difference between Q Comp and non-Q Comp districts suggests that non-adopters would not necessarily provide an appropriate control group (e.g. trends in unobservables might not be similar between Q Comp districts and non-Q-Comp districts) and that I would be better off restricting the sample to only include those districts that adopted the reform during the relevant time period. Thus, the results I present for my student achievement analysis are estimated using only the sample of adopter districts.⁴ Reassuringly, results look similar across the board when estimated using the full sample of all traditional districts in Minnesota (available upon request). This creates a final sample that, for the time period from academic year 2003-2004 through 2009-2010, includes 305 schools and 366,763 student-years. As noted, this base sample corresponds to almost exactly one-third of the K-12 students in Minnesota.

C. School Practice Analysis

To consider whether and how Q Comp changed school practices, I use data from two waves of the Schools and Staffing Survey. Administered every four years, the SASS is a national survey and the largest comprehensive source of data on school organization and staff perceptions. During each round of data collection, the National Center for Educational Statistics

⁴ The school practice analysis using the Schools and Staffing Survey only has two years of data available. Thus, for this analysis, I keep both data from both adopting and non-adopting districts in Minnesota, with the non-adopting districts serving as my control group.

(NCES) survey schools, teachers, and principals, creating national and state-representative data files. The primary sampling unit for the survey is the school, and schools are selected and assigned sampling weights based on sector, location, school-level, and school population. Once chosen, schools provide teacher listings and teachers are similarly stratified and assigned sampling weights based on their subject areas and experience levels. Prior research has used the SASS data to identify trends in teacher qualifications, teacher autonomy, and various labor market outcomes (e.g., Ingersoll 1999, 2006, Liu 2007).

Because the SASS was administered once during the 2003-2004 school year, before Q Comp existed as a program, and once in 2007-2008, it provides an important source of data on school practice with strong construct validity since, unlike other single time-point surveys that have been used to investigate Q Comp, it does not depend on after-the-fact reports about the influence of a particular program. Moreover, although the SASS does not survey teachers in every district in Minnesota, its randomized and weighted design provides a representative sample at the state and district levels. To construct my analysis sample, I keep data from all teachers in districts that were sampled in both 2003-2004 and 2007-2008. Due to power issues, I do not limit to teachers in certain subjects or grades. This sample includes 63 districts in total, with 15 districts that joined Q Comp by 2007-2008. The 2003-2004 data includes 706 teachers and the 2007-2008 data includes 622 teachers.

In the 2003-2004 and 2007-2008 waves, the SASS includes several questions that are particularly relevant to an investigation of Q Comp. As noted, Sojourner et al. (2001) used district-level SASS surveys to determine whether Q Comp districts were more likely to say that they paid teachers based on performance. I use teacher-level surveys to explore teacher pay shifts in greater detail. In particular, I explore the ways that the relationship changed between teacher

reported base pay, bonuses received for extracurricular and leadership activities, and bonuses received as performance incentives shifted in response to Q Comp.

I additionally examine reported changes to teacher professional development and teacher attitudes. While the SASS does not ask teachers to report the total amount of time spent on professional development activities across the year, it does ask teachers to report hours spent in particular kinds of development including professional development around reading instruction, around a teachers' content area, around disciplinary practices, and around computers and technology. These measures allow me to explore the ways that Q Comp changed teacher learning opportunities and to get some idea about whether the intervention resulted in more overall professional development or simply shifted the distribution of professional development across area. The SASS also asks teachers to rate their school atmosphere and their own satisfaction along a variety of dimensions. Several of these are quite relevant to Q Comp's design. About their school, teachers are asked whether there is a cooperative effort among the staff and whether the administration is supportive. About their own work, teachers are asked whether they feel satisfied with being a teacher and whether they feel that they are recognized "for a job well done." I investigate outcomes on each of these survey questions separately to get an idea of how the adoption of whether and in what ways Q Comp changed the teaching climate within program schools.

D. Student Achievement Analysis

To determine whether Q Comp influenced student achievement, I draw on two distinct sources of data, each with its own set of benefits and drawbacks.

The first achievement measure, the Minnesota Comprehensive Assessment is the statewide test used in Minnesota for school accountability measures under NCLB. Test results are publicly available at the grade-school level, however the test presents two challenges for analyzing the effects of Q Comp.

First, the test was rewritten in 2006 and renamed as the MCA-II, and the tests were not designed to be fully equatable. In order to compare outcomes across the different versions of the test, I create normalized average scores within each subject-school-grade by subtracting the yearly average and dividing by the student-level standard deviation of the exam in that year (Jacob, 2007). Specifically, I calculate:

$$z_{sgt} = \frac{y_{sgt} - \bar{y}_{sgt}}{sd_{sgt}}$$

where z is the standardized score and y is the raw score. Sojourner et al. (2011) provide useful evidence that this process allows meaningful comparison between versions of the test by investigating how well a previous year's school-grade test data predicts the subsequent year. They find no significant differences in predictive accuracy between the year that the test changed and any other year.

Another difficulty is that the state test was only given in select grades in its early years. To create a dataset that spans 2004-2010, allowing meaningful estimates of pretrends before the adoption of Q Comp, my MCA sample can only include grades 3, 5, and 7. Thus, although Q Comp is a districtwide adoption that includes all school-grades, my conclusions about the effects on MCA achievement are limited only to these three middle grades.

The second achievement measure I include comes from the Northwest Evaluation Association (NWEA). NWEA tests offer several advantages on the MCA data. NWEA data is available at the student level, allowing me to include student fixed effects in my regressions and potentially control for omitted variable bias that could enter into my MCA results. Also, when it is available, the data spans a larger range of grade-years than the MCA tests, and most districts that use NWEA give these tests twice a year in grades 3-8. While NWEA offers its own method for transforming raw test results into scaled Rasch scores, for comparability purposes I follow the same procedure described above for the MCA tests where I standardize raw spring test scores using yearly student-level averages and state-level standard deviations.⁵

Unfortunately, the NWEA tests are not universally used across Minnesota. Districts choose whether or not to contract out with NWEA and in what year to do so, creating a select sample of districts that is not necessarily generalizable to the whole of the state or of the Q Comp adopter sample.⁶ Table 2.2 compares a set of summary means and standard deviations in 2005 between the full set of regular districts in Minnesota and those that ever used NWEA assessments at some point within the five-year study period. As the table demonstrates, across these observable characteristics, the samples look approximately identical, however the possibility of unobservable differences still makes it useful to examine effects across different samples. Moreover, since districts joined with NWEA in different years across the study period, the table shown underestimates the extent of differences across the samples.

⁵ Note that this method raises a potential issue for comparing program impacts on the MCA and NWEA tests since the test results are standardized using a slightly different set of schools (MCA scores are standardized using all schools in the state whereas NWEA scores are standardized using only schools that work with NWEA). I completed a series of analyses (available upon request) that suggest that these methods do not skew results, comparing means and SDs at the student and school-levels as available across the matched and full sample to see if there were any observable differences. Moreover, when I re-run the tables presented in this paper using matched MCA and NWEA test results that are standardized using the matched sample of schools and districts and using standard deviations at the school-grade level (since student-level standard deviations are only available for the entire state), results are qualitatively similar.

⁶ The NWEA sample does not include any charter schools.

Another concern around the use of NWEA data is that districts have considerable freedom over when they administer NWEA tests so that testing dates vary considerably at the district level, unlike the MCA test, whose testing days are standardized across the state. Following Springer, Ballou, and Peng (2008) who also use NWEA test data, I include controls for testing date in all of my regressions that use NWEA assessments as the outcome.

Taking the limitations of these different data sources into account, I define four samples and estimate impacts separately for each one. The first, which I refer to as the MCA full sample, includes all regular schools across the state of Minnesota, but as mentioned previously is limited to grades 3, 5, and 7. The second, which I refer to as the NWEA full sample includes grades 3-8 but only the districts that offer the NWEA assessment. Finally, I create a matched set of samples of grades and schools that administered both tests for each sample year. Thus, the MCA and NWEA matched samples are made up only of the district-years that work with NWEA and include only grades 3, 5, and 7 for which both tests were available during all study years. All samples span academic years 2003-2004 and 2009-2010.

While both the MCA and NWEA assessment have their individual challenges as detailed above, there are several benefits that come from using both tests simultaneously to explore Q Comp's causal impacts. First, the differently written tests might pick up different constructs, and it is useful to see whether gains on one test are reflected by gains on the other. Second, the two test results are used quite differently by schools and districts, with MCA providing high-stakes results used by the state for school accountability purposes and NWEA providing formative information meant to inform teachers' instructional decisions. These differences in use could potentially generate differential effects. Third, many districts that used NWEA set their Q Comp goals based upon this test rather than the MCA test, raising the possibility that there might be

effects on one of the tests that would not be detectable on the other. I investigate this possibility in more detail in section V(ii).

For an initial comparison between the two assessments, Figure 2.2 shows trends over time in test results for the matched samples of MCA and NWEA data, broken into cohorts of Q Comp adopters. While the shape of the graphs is not identical, the trends look quite similar between the MCA and the NWEA tests. For instance, looking at the math results, the 2007 Q Comp adopters begin at the highest level in 2004, and exhibit a small test drop in 2006 before returning to the original averages on both MCA and NWEA tests. Similarly, 2008 adopters display a steady increase in scores through 2009 before dropping off slightly. These graphs also demonstrate the significant and constant difference in scores and trends between Q Comp adopters and non-adopters, validating the choice described in a previous section to focus analysis on only the sample of adopters. The graphs also provide some evidence as to the lack of obvious Q Comp impacts on student achievement, since there is little indication that the different adopter cohorts demonstrate a clear increase in test averages following the year of Q Comp adoption.

V. RESULTS

I present results in three sections. In the first section, I look at how Q Comp changed school practices for evidence about implementation. In particular, I look to see whether the program created expected changes to teacher salary structures, teacher professional development, and school climate. In the next section, I look at the impact of the program on student achievement. I begin by looking at overall effects. Then I look to see whether effects were concentrated within schools' targeted areas of improvement. Lastly, I use findings from a qualitative study of Q Comp in three districts as the basis for coding programs according to

district-level differences in design, and I examine interactions between Q Comp implementation and district program type.

i. Did Q Comp Change School Practices?

Tables 2.3-2.5 present weighted least square estimates on several SASS outcomes using the DD model detailed in Equation (1). The model looks at teacher-level differences in Q Comp districts pre and post-program adoption as compared with teacher-level differences in other Minnesota school districts surveyed by both the 2003-2004 and 2007-2008 waves of the SASS. All specifications include district fixed effects and controls for characteristics of the school's student body. Results are weighted to reflect the effect on the average teacher using NCES-provided teacher weights.

A. Teacher Pay

Table 2.3 focuses on changes in teacher pay, measured in 2011 dollars. The first three columns of the table include only school-level controls while columns four through six add dummies for years of teaching experience, masters degrees, and the interaction between experience and degree in order to control for compositional changes to the workforce that might affect average salaries. Columns one and four represent the change in teachers' base salary amount that can be attributed to the adoption of Q Comp, columns two and five show the change in performance bonus amount, and columns three and six show the change in additional salary received for other school activities (e.g. extracurricular leadership, etc.).

According to the results in column one, the addition of Q Comp to a district lowered teacher base salaries by around \$3000. When I control for the interactions between years of

teaching experience and higher education degrees in column four, however, the results decrease considerably and lose significance, suggesting that the base salary change derives from a change in the composition of the teacher workforce in Q Comp districts toward less credentialed teachers rather than a change in teacher salary schedules.

In contrast, I find in column two that teachers received a significant increase of around \$750 in the category defined by the SASS as “income from other school sources, such as a merit pay bonus, state supplement, etc” and that this estimate stays almost exactly constant even when I add teacher-level controls in column five. The size of this pay increase that I identify makes sense within the structure of district programs, which typically offered teachers the opportunity to earn a maximum of one to two thousand dollars in performance bonuses. Thus, as intended, Q Comp does appear to have shifted teacher pay toward what teachers report is a more performance-based system. Yet the size of this shift and especially the size of the bonuses were relatively small, with the \$750 increase in performance bonuses representing around one percent of teachers’ total pay for the year.

Columns three and six estimate changes in any additional teacher salary components including “extracurricular or additional activities such as coaching, student activity sponsorship, or teaching evening classes.” These estimates are small and insignificant both without and with the addition of teacher-level controls, suggesting that teachers did not see a decrease in other hidden pay categories to offset the additional bonus money. At the same time, as noted previously, districts seem to have offset the additional amount that they were paying teachers in salary bonuses through shifts in the teacher workforce.

B. Professional Development

Because the SASS does not ask teachers about all their professional development experiences but only about professional development in certain categories, the evidence that this data has to offer about overall changes to teacher professional development caused by Q Comp is somewhat limited. However, the results presented in Table 2.4 are informative and again consistent with qualitative reports about Q Comp (Schwartz, 2012).

Teachers in Q Comp districts saw a significant increase of around 3 extra hours per year of professional development focused on reading instruction. While 3 hours per year might seem relatively small, the increase builds on an initial average of around 7 total hours of reading professional development, meaning that the increase increased reading professional development time by nearly 50 percent compared with teachers in other districts across Minnesota. At the same time, there were no significant changes in the amounts of professional development in other areas. This makes sense given that Q Comp districts had to choose to focus teacher learning exercises on either reading or mathematics instruction and the SASS does not specifically ask teachers about professional development in mathematics. To try to determine whether or not Q Comp districts also added significant quantities of mathematics professional development, I tried restricting my sample to only math teachers and focused on the column 2 model specifying professional development in the teacher's content area. At this point, however, the sample sizes became small enough that I would not have been able to detect effects as small as those identified in reading.

Alongside the calculation of the number of hours teachers devoted to various types of professional development, there is the more important question of how useful these different development opportunities were to the teachers who participated. A central premise of Q Comp is that, through the action of site-based teacher-leaders, professional development would become

more relevant and more instrumental in increasing student learning. Importantly, teachers' reports on the SASS provide no evidence that professional development became more useful on average. Panel B of Table 2.4 shows that teachers' ratings about the usefulness of different types of professional development remained effectively unchanged due to the adoption of Q Comp in their districts. Since teacher learning represented a major programmatic goal, this finding seems quite problematic for the program as a whole. As the next section demonstrates, however, teachers did report other changes in their schools due to the program's adoption.

C. Teaching Climate

Table 2.5 shows changes to teachers' responses to a variety of questions about the teaching climate in their schools and their professional satisfaction due to the adoption of Q Comp. Strikingly, Q Comp significantly shifted teachers' willingness to agree with the statement that "staff members are recognized for a job well done" raising responses by around 0.35 points on a five-point scale. There are several interpretations of this point. One possibility is that this change is a direct result of the pay incentives offered by the program, which offered teachers performance-based rewards even if they did not result in higher overall salaries. Another is that the changes in teachers' perceptions of their school's climate have less to do with shifts in compensation and more to do with the other elements of the program. For instance, if Q Comp helped to create a more professional atmosphere where teachers were focused on clear goals and felt like they were being acknowledged for their work in meeting these goals, they might feel like they were being recognized regardless of their overall pay.

While there is no way to use this data to identify the exact mechanism by which Q Comp caused teachers to feel more recognized, we can see an additional result in the data that the

adoption of Q Comp made teachers significantly more likely to find the school administration to be “supportive and encouraging.” Since Q Comp is directly focused on teacher leadership and makes little provision for administrator involvement, this is an unexpected result. Again, it might simply be a change in teachers’ perceptions or even a Hawthorne effect caused by entry into the program but, coupled with the increase in reading professional development, it might also mean that Q Comp was altering school climates in significant ways.

At the same time, there was no significant change in teachers’ reports about the level of staff collaboration taking place in their schools. After the adoption of Q Comp, teachers were no more likely to agree with the statement that “there is a great deal of cooperative effort among the staff members” or that most colleagues share beliefs and values than they were before the advent of the program. Moreover, even though teachers felt more recognized for their work and better supported by their administration because of Q Comp, they were no more or less satisfied with their work than they had been before the program’s adoption.

ii. Did Q Comp Change Student Achievement?

Tables 2.6-2.10 and 2.12 present weighted least squares estimates of Q Comp’s effects on MCA and NWEA testing outcomes using the models from Equations 1-3. As noted, these models compare pre- and post-adoption differences in districts that adopted Q Comp with changes in districts that had not yet adopted Q Comp during that time period. Each regression controls for student characteristics and includes either grade-school and grade-year fixed effects (MCA results) or student fixed effects (NWEA results). Results are weighted to reflect the effect on the average student.

A. Main Effect

Table 2.6 presents Q Comp effects on student achievement as measured by both the Minnesota Comprehensive Assessment statewide tests as well as the Northwest Evaluation Association's assessments used by a subsample of districts. For each separate outcome/sample, column 1 presents the simple DD specification described in equation 1. Column 2 adds in a dummy variable that turns on for observations that occur two or more years before the district's adoption of Q Comp, as described in equation 2. Column 3 adds in a control for district-level pretrends in achievement, as described in equation 3. Accordingly, columns 1 and 2 for the MCA full sample follow on results presented by Sojourner et al. (2012).⁷ Regressions for the MCA samples were run at the grade-school level and thus include grade-school and grade-year fixed effects, while regressions for the NWEA samples were run at the student level with individual student fixed effects. Each set of regressions controls for a variety of demographics at the grade-school level.⁸ Results are weighted at the student level.

As the table shows, Q Comp's overall effects on student achievement appear to have been relatively small to non-existent. On both MCA math and reading results in the full sample, I do not detect significant effects and the point estimates are less than 0.025 student-level standard deviations. Even within potential bounds for the estimate suggested by the standard errors, these effects would be small at best. As a point of comparison, Krueger (2003) found that moving students to small classes in elementary school raised test scores by around 0.2 standard deviations. Importantly, while the significant negative results on the Q_2ormore dummy in

⁷ The most recent draft of Sojourner et al. takes advantage of student-level state testing data that I do not have access to. The additional power from this data means that the authors find some significant effects that look insignificant in my results (specifically, they find a significant effect of around 0.03 SDs on overall reading achievement). While the significant findings are an important addition, the small effect size does not change the conclusion that Q Comp's effects on achievement have been minor at best.

⁸ Leaving out these demographic controls does not significantly change the final estimates.

column 2 suggest that math scores were already rising in Q Comp districts prior to the adoption of the program, the addition of district-level pretrends does not diminish the point estimates and indeed even raises the value slightly in some cases. Moreover, results on MCA math scores look relatively similar although slightly larger when confined to only the sample of schools for which I have matched NWEA data.

Q Comp's impacts on math achievement as measured by NWEA track the MCA results quite closely. Estimates are slightly larger in the NWEA matched sample (columns 4-6 of Panel B) that restricts the NWEA data to only grades 3, 5, and 7 rather than including grades 3-8 (columns 1-3 of Panel B), but again most estimates are insignificant. It is useful to note however that MCA and NWEA results track quite closely, suggesting that the matched NWEA data provides a fairly close approximation of the larger samples. When looking at the NWEA results, it is also important to note that the sample size drops considerably in the models for which I include a control for district-level pretrends. This is because several districts began using NWEA tests either in the same year that they joined Q Comp or during the previous year. It is impossible to estimate pretrends for these districts since they do not have two or more years of data from before Q Comp. Moreover, it is difficult to know a priori how these districts might skew results if NWEA adoption was linked to Q Comp adoption. Reassuringly, the fact that the results with controls for pretrends in column 3 of the matched sample results look relatively similar to the other columns suggest that this concern is unfounded.

Results for reading achievement as measured by the MCA remain slightly positive, however as with the math results, it is not possible to say at a reasonable level of confidence that the impacts are greater than zero. As with the math results, the addition of controls for district pretrends does little to change the overall estimates on the Q Comp effect. Results for NWEA

reading achievement track MCA results less closely than they did for math, yet the values on each outcome fall within the others' margin of error.

B. Subgroup and Distribution Effects

The estimates shown in the previous sections show that Q Comp has had a limited overall effect, but they say nothing about how achievement levels might have shifted among students. Table 2.7 looks at effects on students at different points in the achievement distribution. On the NWEA tests, where I have access to student-level data, I focus only on the group of students who were present in the year before Q Comp adoption. I divide these students into achievement quartiles based on their test results in that pre-adoption year and then run separate regressions for each quartile to estimate the impact of the Q Comp program on these particular students. On the MCA tests, where I only have school-grade level results, I partition the data by looking individually at student-grades where state-defined student proficiency rates were less than 50 percent in the pre-adoption year and student-grades where state-defined proficiency rates were greater than 85 percent in the pre-adoption year.⁹ To simplify results, I present results using only the full samples available for each test and only model specifications (1) and (2).

These results are striking, showing across both tests that Q Comp has had a large and significantly positive impact on the students who began lower on the achievement distribution and a corresponding negative impact on the students who began higher on the distribution. The shape of this trend is clearest across the NWEA quartiles, where students whose test scores placed them in the lowest quartile in the year before Q Comp adoption averaged gains of about

⁹ In math, approximately one-quarter of the student-grades in the adopter sample had <50 percent of students proficient and approximately one-quarter of the student-grades had >85 percent of students proficient. In reading, approximately seven percent of school-grades had <50 percent of students proficient and 35 percent of school-grades had >85 percent of students proficient.

0.25 standard deviations in math and 0.3 standard deviations in reading whereas the students who began in the highest testing quartile lost around 0.15 standard deviations in math and 0.25 standard deviations in reading. The same trend can be traced through the middle two quartiles. Similarly, in the MCA data, we see a significant positive effect of Q Comp for student-grades where most of the students were not rated as proficient before program adoption and a negative effect for student-grades where more than 85 percent of students were initially rated as proficient.¹⁰

One threat to the validity of such estimates that divide students into groups based on initial values of the dependent outcome is that they might pick up regression to the mean rather than a genuine effect. Several analyses can help tease out these differences. One possibility is to change the dependent outcome to be the share of students within a fixed state distribution rather than the actual achievement score. While such analyses that might help to verify these results were not possible within the time frame of this dissertation, they are crucial next steps in this program of research.

Assuming the results outlined above continue to hold, do they mean that Q Comp's heterogeneous effects are confined to achievement level alone or do these effects also differ across other student characteristics? To test this, Table 2.8 explores Q Comp's effects across commonly defined student subgroups. Using the MCA data, I divide estimates by gender and race, and also estimate effects separately for students who qualified for free or reduced price lunch. Note however that school-grades with fewer than 10 students of a particular subgroup were not included in the estimates due to data restrictions. Using the NWEA data, I only have information on student gender and race and thus only test for differences across these subgroups.

¹⁰ These results remain similar when I include Q Comp non-adopters in the regressions.

Interestingly, there is no evidence in Table 2.8 of meaningful differences in effect across gender, race, or SES subgroups. Indeed, in contrast to Table 2.7, Table 2.8 looks nearly identical to the table of effects for the full sample, with estimates of effects for male and female students, white and non-white students, and free lunch students all hovering around zero.

C. Subject and Test Focus

As part of the program design, individual schools in Q Comp districts are required to put together a set of yearly testing goals that then provide the basis for several of the Q Comp elements, including the distribution of school-level performance bonuses and the emphasis of professional development activities. When setting these goals, schools can choose whether to focus on math or reading as well as whether to set goals based on MCA or NWEA test measurements. Although slight, the differences across subject and test outcome estimates in the previous two sections raise an important set of questions about whether these focus areas have affected program effects. One possibility is that the overall program impacts shown in Table 2.6 are misleading because they do not hone in on test impacts in particular school focus areas. This section explores the interactions between Q Comp adoption and yearly subject/test focus.

Yearly school performance goals play a potentially important role in Q Comp operations. In nearly every Q Comp district, meeting these goals leads to the distribution of teacher bonuses for schoolwide performance gains. These goals also serve to organize the professional development elements of Q Comp by providing the focus area for professional learning communities, peer coaching initiatives, and student data analysis. School goals must be approved by MDE at the beginning of each year, and those that are deemed insufficient because they are too small or not specific enough are returned for revision. Figure 2.3 shows the distribution of

goals at the school-year level between 2005 and 2010. As the figure demonstrates, many more schools choose to focus Q Comp efforts on reading than math. Similarly, even of those schools that contract with NWEA, many more schools tend to set achievement goals based on NWEA benchmarks (Figure 2.4).

Interpreting any significant interactions between Q Comp and a particular focus area would not be particularly straightforward. The difficulty here is that gains limited to particular focus area might represent genuine improvements in student learning that accrued because Q Comp teachers and administrators were concentrating their efforts on these targets. Alternatively, the gains might represent teachers “gaming the system,” setting yearly goals that they knew they could meet in order to receive the final payoff. Note however that the average payoff for these school-level bonuses is quite low, averaging less than \$350 per year.

In either case, the issue is moot since there is little evidence that program gains meaningfully track school focus areas. Tables 2.9 and 2.10 present saturated models of interactions between Q Comp and either subject- or test-focus areas run without an included constant. Note that the signs on the coefficients in Table 2.9 do point in the expected directions. That is, Q Comp impacts on math achievement lean in the positive direction for schools that set math improvement goals and in the negative direction for schools that set reading achievement goals, and the signs flip with respect to reading achievement.¹¹ Nevertheless, the central point is that none of these effects are statistically distinguishable from zero, regardless of sample or specification. The same is true when test focus is examined in Table 2.10. Again, there is no

¹¹ One might hypothesize that the relationship between yearly goals and achievement could be more complex than these initial specifications allow. For instance, maybe a school makes gains in an area based on the number of cumulative years that it has set goals in a particular subject area rather than during the single year of targeted improvement. However, tests of alternate specifications like these yield no new significant findings (results available upon request).

significant evidence that schools post better outcomes on the MCA test when they set MCA rather than NWEA goals or vice versa.

These results suggest that any impacts that Q Comp might have had on student achievement have not been hidden by specifications that consider only overall effects rather than targeted effects in targeted years. The results provide further evidence that Q Comp's overall effects have been highly limited at best.

D. Program Design

Qualitative work presented later in this dissertation (Schwartz, 2012) offers a hypothesis about how and why Q Comp effects might differ by district. In case studies of three districts, I find that district programs looked quite similar along many dimensions, largely as a result of the Minnesota Department of Education's efforts to structure and standardize the program. For instance, although there is some variation in the amount of bonus pay offered for various categories of performance benchmarks, this variation averages only several hundred dollars or less than 1% of teachers' annual salaries. Moreover, nearly every district awards bonus pay for similar categories of benchmark. Similarly, while districts did adopt marginally different systems for their teacher observations and evaluations, most districts chose similar sorts of rubrics (either Charlotte Danielson or a district-produced version that resembled the Danielson rubric) and adopted similar procedures with regard to the number of observers present during an observation and the number of observations per year.

However, the state was least active in the area of professional development and teacher learning. During both the initial application process and the later implementation, districts were given considerable leeway over the ways that teacher-leaders would interact with the rest of the

district and over the types of professional development opportunities that the program offered. This could potentially represent a crucial piece of the teacher improvement system. As noted, one reason that these systems might make a difference to student improvement is because they could shift teachers' incentives while simultaneously helping teachers acquire the skills to meet these incentives. As such, the ways that districts set up the system to help teachers acquire these skills could be a crucial factor in programmatic success.

Case study work suggested three different ways that districts were thinking about improving teachers' skills and knowledge. In one type of district, *other teachers* were the source of new expertise and the district design made peer coaching, frequent peer observation, and best-practice sharing the central focus of the Q Comp program. In other districts, *student data* was meant to provide the basis for teacher learning, and the Q Comp programs emphasized data disaggregation and teacher-led data analysis. In the third set of districts, teachers were meant to learn from *external sources*, and these districts tended to partner with external professional development programs or bring in new sources of outside research to supplement the current teacher knowledge base. In the article describing these district types, I hypothesized that the third program type appeared strongest from a design point of view, since it created a clear conduit for teachers to learn new skills and knowledge to supplement their current practice. This seemed like a necessary element for a teacher improvement program that hoped to raise teachers' capabilities to meet various incentives. In contrast, the other program types offered teacher incentives but did not provide a clear source for new expertise.

To test this hypothesis, I relied on the information provided in district applications to the Minnesota Department of Education for initial Q Comp funding. These applications include information about the overall operation of the district's Q Comp program, lists of all teacher

career ladder positions added by the program, budget worksheets, school-level goals, and descriptions of how teacher professional development is designed to support these goals. Using the characteristics provided by my case studies, I assigned Q Comp districts to one of the three program categories. Figure 2.5 shows the distribution of districts across these program types. Note however that this is not the ideal coding setup. District's Q Comp applications are written before program implementation and describe the program as it was meant to operate, not necessarily as it actually did operate. Thus, some districts might have designed a program to emphasize one type of teacher learning and created a program that actually emphasized another. Indeed, during case study work with three districts, I found several areas where district programs differed in meaningful ways from the proposals submitted in the initial applications.

Moreover, as a source of causal inference, this analysis raises several concerns that go beyond those when considering the effects of Q Comp alone. Districts were not randomized into program type, and thus program types are likely to be collinear with other district characteristics. Table 2.11 shows summary statistics for Q Comp districts, separated by their district-level program types. As is clear, the districts differ in several areas alongside their program designs. Districts that adopt data focused Q Comp designs tend to average lower enrollment than those that emphasize learning from other teachers or external sources. The districts that emphasize learning from external sources tend to be those with the greatest percentage of disadvantaged students. At the same time, these districts also pay higher teacher salaries and employ more teachers with masters degrees. These differences are to be expected; case study work suggested that the design approach that districts adopted with respect to teacher professional development depended in large part upon the activities that were taking place in the district before the state program existed. At the same time, the differences make differential effects more difficult to

interpret. While my regressions control for many of the variables in the summary tables, there is likely to be other unobservable variation as well. A positive effect on the Q Comp*district-type coefficient might indicate a causal relationship between design type and achievement, but the observed effect might also have been driven by another omitted variable.

Nonetheless, the findings from such interacted regressions are still of interest since, if significant, they can point the way toward future study. Moreover, these results offer a useful initial check on the qualitative results in Schwartz (2012).

Table 2.12 provides regression results from specifications that group Q Comp districts into the three discrete categories described above. As with the subject and test focus models in the previous section, each category is included in the regression and regressions are run without an included constant. Panel A, columns 1-3 show the association between the different program types and student math achievement on the MCA test for the full sample of Minnesota schools. In this panel, we can see a significant increase in test scores only among districts that adopted externally focused Q Comp programs. This effect stays constant across the three specifications. There is no evidence of a similar effect on reading scores, a finding that aligns with previous work suggesting that external professional development tends to generate greater gains in the realm of mathematics than reading.

However, the evidence of an effect in either math or reading among districts with Q Comp programs emphasizing teacher learning from external sources disappears when we move to the sample of schools that also use NWEA tests. This suggests that the effect is only taking place in the subsample of districts that did not contract with NWEA. Results in Panel B that use NWEA test results support this conclusion since there is no evidence of a positive effect among the subsample of external learning districts. Interestingly, however, using NWEA tests as the

outcome, the districts that focused on student data see the greatest gains from joining Q Comp. This is not an entirely implausible result; unlike MCA tests, NWEA tests are given at the beginning and end of the year and are designed to be used by teachers to inform classroom instruction. It seems possible that districts that both contracted with NWEA and emphasized the analysis of student data in their Q Comp plan made gains that others did not. However, the fact that these gains do not show up for the same schools and students when the MCA test is used as the outcome rather than the NWEA test makes the finding less substantial, and suggests that these results might simply be due to chance rather than a true effect.

VI. CONCLUSION

This paper uses the staggered adoption of Q Comp by districts in Minnesota as a means to estimate the impact of program adoption on several outcomes, including teacher salaries, professional development and teaching climate within districts, and student scores. I find that Q Comp had clearly discernible effects on the way that teachers were getting paid, as well as their exposure to certain kinds of professional development, and their opinions of their school climate. When I looked for similar impacts on student test scores, I found little evidence of corollary impacts on average achievement. Yet the program appears to have had pronounced effects on the achievement distribution, raising scores for the lowest achievers and lowering scores for the highest achievers.

Why might we see such effects from this kind of program and what are their policy implications? Several studies of educational interventions have noted differential effects by prior student achievement, with the direction of this differential dependent on the particular intervention. A recent study of an algebra-for-all intervention program in Chicago, for instance,

found that the program had negative effects on the test scores of initially high achieving students (Nomi, forthcoming). In contrast, a study of the effects of a KIPP charter school using a randomized sample of students drawn from the school's entrance lottery found that the school's effects were significantly higher for students who entered with low test scores (Angrist et al., 2010). In this case, there are several reasons why the effects might have been positive for students on the lower end of the achievement spectrum and negative for students on the higher end of the achievement spectrum. One possibility is that teachers systematically shifted their attention to particular students in their classes due to the intervention. With new incentives and more feedback, teachers might have become more aware of the need to cater to their low achievers. Another is that teachers actually changed the way they taught in ways that served certain students better than others. Teachers in Q Comp districts received new kinds of coaching and professional development driven through peer-to-peer interaction. It is possible that the practices that teachers shared with each other tended to offer more to low than high achievers. A third possibility is that the effects were driven by decisions at the school or district level. Under Q Comp's design, schools and districts had considerable latitude to determine various focus points for their program. Perhaps these units aimed the program at low achieving students in ways that were not immediately observable in the district program data that I had available. For instance, when schools set yearly testing goals, they might have emphasized low achieving areas. While I did test to see whether schools appeared to make the greatest gains in their subject focus areas, these tests did not take into account the size of different goals and would not necessarily have uncovered differential goals for differently achieving populations.

If verified through further analysis, Q Comp's differential effects on low and high achievers provide evidence that the program has had meaningful impacts on classrooms and

students across the state of Minnesota. The direction of these effects suggest that the program might over time narrow gaps in achievement across the state, although it is troubling that this would occur at the expense of higher achieving students. Equally troubling is the fact that Q Comp has done little to raise average scores across the group of adopting schools as a whole.

Q Comp, with its combination of low levels of district funding, wide-ranging but weakly specified requirements over teacher pay, evaluation, and professional development, and slim instructional guidance bears close resemblance to the developing contour of teaching reform across the United States. Programs like Q Comp are set to launch in a host of states over the next several years. Many of these reforms will reach even farther than Q Comp in that they will be instituted as state-wide requirements rather than opt-in programs. However, the basic structure of these systems will still be quite similar to the Minnesota version. This paper provides evidence that such programs are by no means certain to advance states' school improvement goals.

At the same time, there are several important limits to the generalizability of the conclusions that I find here. First, the student achievement analyses only pertain to the select subjects (math and reading) and grades (3, 5, and 7 on the MCA test and 3-8 on the NWEA test). With respect to subject, it is reasonable to expect that Q Comp's effects might have been concentrated in math and reading since these are the subjects emphasized by the program content. However, the intervention extends across all grades within a district, so it would have been helpful to have access to a larger grade range of testing data and to be able to test for differential effects across grade levels. One could imagine that Q Comp would have quite different effects at the elementary level where most teachers teach all subjects in self-contained classrooms than at the high school level where teachers specialize in single subjects.

Another significant limit on the generalizability of this study stems from the fact that Q Comp is an opt-in program that has mostly drawn districts with more resources, higher initial test scores, and higher population density than the typical Minnesota district. Judging how such a program might have looked if it were implemented statewide as most of the Race to the Top plans will be is difficult. On the one hand, the districts that took on Q Comp might be expected to be the most capable districts in the state for designing and implementing complex program designs. Additionally, by joining the program, the districts demonstrated that they at least partly bought into the program's goals. In this case, it seems quite problematic that the program did not produce greater overall achievement effects in what appeared to be a reform-conducive environment. On the other hand, since the districts that opted into the program were quite well-resourced even before joining Q Comp, they might have already implemented many of the teacher supports that Q Comp was meant to create. In this scenario, Q Comp might have had greater effects if it had instead been focused on the districts with the greatest need.

Moreover, there is still much to learn about whether different but related program designs might have been more successful. One possible explanation for the lack of overall effects was that the high levels of district autonomy in the Q Comp design structure created substantial variation across district adopters. These differences were difficult to sort out within the bounds of this study, in part because the adoption mechanism limits my ability to make causal claims, and in part because information about district designs was restricted to the information that districts sent to MDE before implementation took place. Within these constraints, I found slight evidence that districts that focused on bringing in outside expertise about teaching as part of their Q Comp design made larger gains than districts that emphasized best-practice sharing or data analysis but relied on internal capacity to bring these things about. This finding might provide

some justification for future state-level program designs that provide more active support in the realm of instructional learning. In a later chapter of this dissertation, I argue that a major flaw in Q Comp as designed at the state level was the lack of state capacity to provide support and training around teaching improvement for districts engaged in the program. It is possible that the districts that focused on bringing in external instructional expertise were finding substitutes for this gap in the state-level program. However, as noted, the ability to draw firm conclusions about district-level program variation from the quantitative data was quite limited.

Given the uncertainty of these conclusions, there is considerable need for further research into both the operation and outcomes of current teacher improvement systems. Purposeful piloting by states of different models of such systems and/or different levels of district autonomy might help to illuminate the possibilities for coordinated state and district action and the most important sources of district-to-district variation. Similarly, such small-scale experiments would provide further evidence as to whether these programs can be tailored to create more consistent success or whether states would do better to expend their teacher and school improvement resources elsewhere.

REFERENCES

- Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., and Walters, C. R. (2010). "Inputs and impacts in charter schools: KIPP Lynn." *American Economic Review: Papers and Proceedings*. 100: 1-5.
- Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Measures of Effective Teaching Project, Policy Brief.
- Borman, G., Slavin, R.E., Cheung, A., Chamberlain, A., Madden, N.A., & Chambers, B. (2007). "Final reading outcomes of the national randomized field trial of Success for All." *American Educational Research Journal*. 44(3): 701-731.
- Camerer, C.F., and Hogarth, R.M. (1999). "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty*. 19(1-3): 7-42.
- Coburn, Cynthia. (2001). "Collective sensemaking about reading: How teachers mediate reading policy in their professional communities." *Educational Evaluation and Policy Analysis*. 23(2): 145-170.
- Cohen, D.K. and Spillane, J. (1992). "Policy and practice: The relations between governance and instruction." In G. Grant (Ed.), *Review of Research in Education*. 18: 3-49.
- Cohen, D. K. & Moffitt, Susan, L. (2009). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Press.
- Correnti, Richard and Brian Rowan. 2007. "Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs." *American Educational Research Journal*. 44(2): 298-338.
- Dynarski, S. (2004). "The new merit aid." Pp. 63-97 in *College choices: The economics of where to go, when to go, and how to pay for it*. Ed. Caroline M. Hoxby. Chicago: The University of Chicago Press.
- Dynarski, S. (2008). "Building the stock of college educated labor." *The Journal of Human Resources*. 43(3): 576-610.
- Fryer, R. G. (2011). "Teacher incentives and student achievement: Evidence from New York City." NBER Working Paper No. 16850.
- Glazerman, S., and Seifullah, A. (2010). "An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year Two Impact Report." Washington: Mathematica Policy Research. May 17.

- Glazerman, Steve, Allison McKie, and Nancy Carey. (2009). "An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year One Impact Report" (Washington: Mathematica Policy Research, Inc.).
- Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., Jacobus, M. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study*. (NCEE 2010-4027). Washington DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Ingersoll, R. M. (2006). *Who controls teachers' work?: Power and accountability in America's schools*. Cambridge: Harvard University Press.
- Ingersoll, R. M. (1999). "The problem of underqualified teachers in American secondary schools." *Educational Researcher*. 28(2): 26-37.
- Krueger, A. B. (2003). "Economic considerations and class size." *Economic Journal*. 113: F34-F63.
- Liu, X. (2007). "The effect of teacher influence at school on first-year teacher attrition: A multilevel analysis of the Schools and Staffing Survey for 1999-2000." *Educational Research and Evaluation: An International Journal on Theory and Practice*. 13(1): 1-16.
- Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., Peng, A. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report*. Rand Corporation.
- Miles, K. H., Odden, A., Fermanich, M., & Archibald, S. (2004). "Inside the black box of school district spending on professional development: Lessons from five urban districts." *Journal of Education Finance*. 30(1): 1-26.
- National Center on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Accessed online at http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf on June 19, 2012.
- Newmann, B., Smith, B.A., Allensworth, E., & Bryk, A.S. (2002). "Instructional program coherence: What it is and why it should guide school improvement policy." *Educational Evaluation and Policy Analysis*. 23(4): 297-321.
- Nomi, T. (forthcoming). "The unintended consequences of an algebra-for-all policy on high-skill students: Effects on instructional organization and students' academic outcomes." *Educational Evaluation and Policy Analysis*.
- Rosenberg, S. *Organizing for quality in education: Individual and systemic approaches to teacher quality*. University of Michigan Dissertation Essay.

- Sawchuk, S. (2009). TAP: More than performance pay. *Education Week*, April 1. 28(27): 25–27.
- Siemson, E., Roth, A.V., Sridhar, B. (2008). “How Motivation, Opportunity, and Ability Drive Knowledge-Sharing: The Constraining Factor Model.” *Journal of Operations Management*. 26: 426-445.
- Sojourner, A., West, K., Mykerezi, E. (2012). “Teacher pay reform can work: Panel data evidence from adoptions of Q Comp in Minnesota.” Working Paper, University of Minnesota.
- Solomon, Lewis, J. Todd White, Donna Cohen, and Deborah Woo. (2007). “The Effectiveness of the Teacher Advancement Program.” Santa Monica, CA: National Institute on Effectiveness in Teaching, April 2007.
- Spillane, J. P., and Jennings, N. E. (1997). “Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective.” *Teachers College Record*. 98(3): 439-481.
- Springer, M., Ballou, D., and Peng, A. (2008). “Impact of the Teacher Advancement Program on Student Test Score Gains: Findings from an Independent Appraisal.” (Working Paper 2008-19). National Center on Performance Incentives. Vanderbilt University.
- Springer, M.G., Hamilton, L., McCaffrey, D.F., Ballou, D., Le, V., Pepper, M., Lockwood, J.R., and Stecher, B.M. (2010). *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives for Teaching*. National Center on Performance Incentives. Vanderbilt University. March 21.
- Springer, M. G., Pane, J. F., Le, V., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., Stecher, B. (2011). “Team pay for performance: Experimental evidence from the Round Rock Pilot Project on Team Incentives.” *Educational Evaluation and Policy Analysis*. May 14, 2012.

Figure 2.1: Number of Q Comp Districts by Year

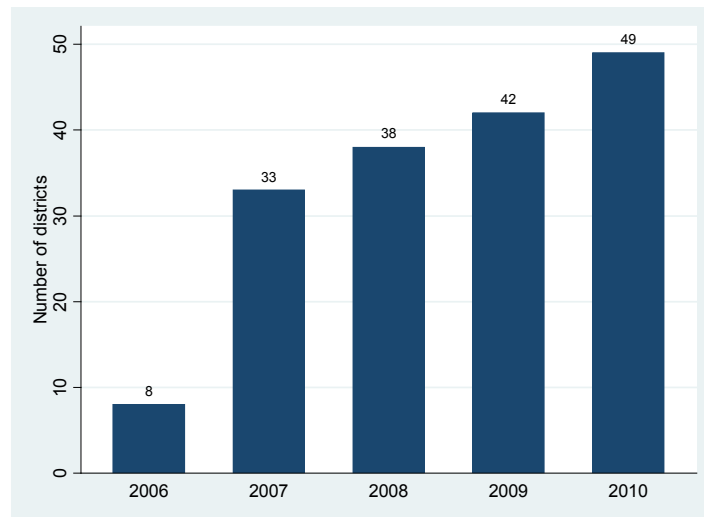
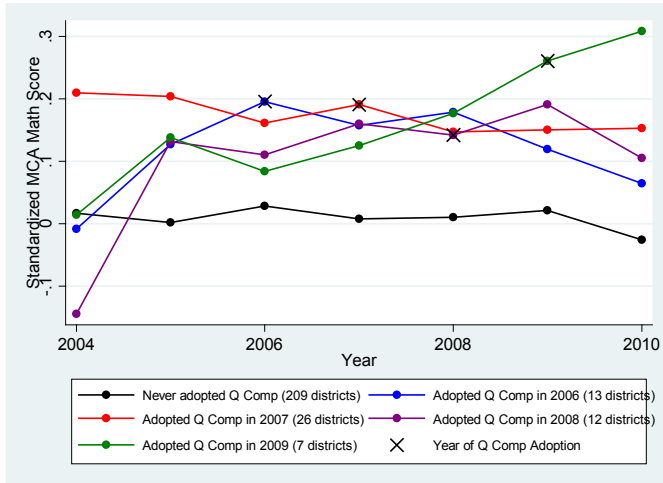
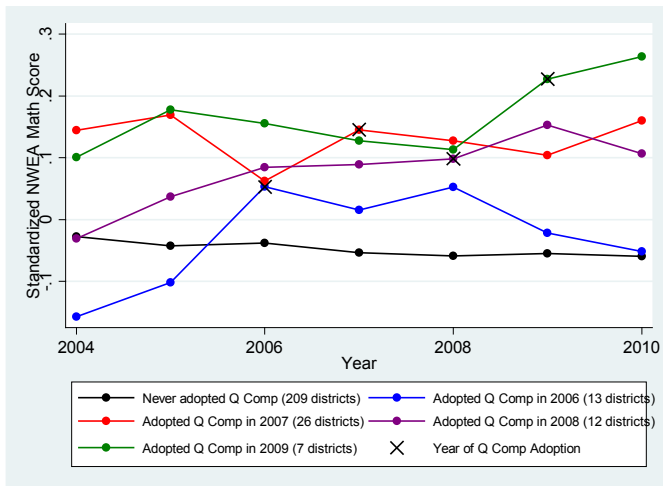


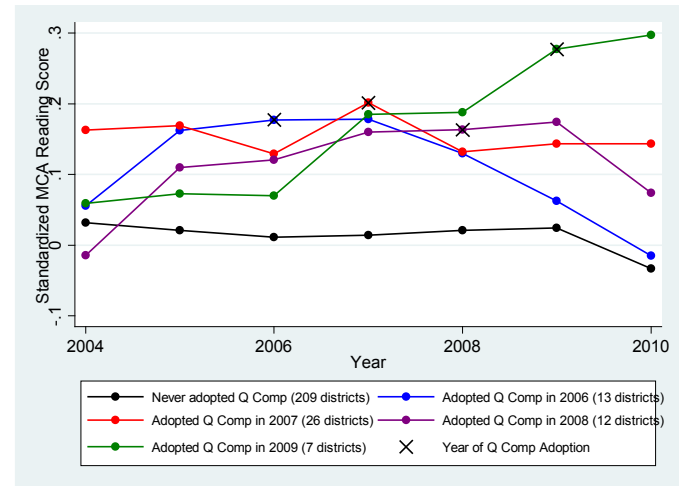
Figure 2.2: MCA and NWEA Test Trends by Q Comp Adoption Cohort



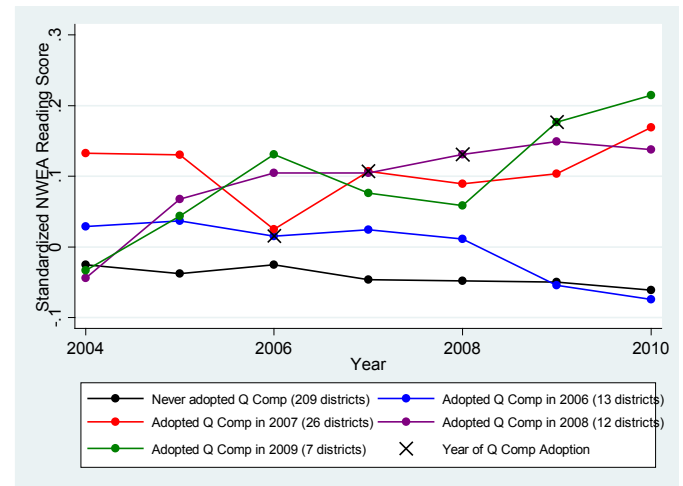
(a) MCA Math Scores



(b) NWEA Math Scores



(c) MCA Reading Scores



(d) NWEA Reading Score

Notes: Samples composed of school-grade-years for which both MCA and NWEA scores were available. Grades shown include 3, 5, and 7. Test scores standardized to have a mean of 0 and standard deviation of 1 within subject-year using student-level test means and standard deviations.

Figure 2.3: Subject Focus Distribution (School*Years)

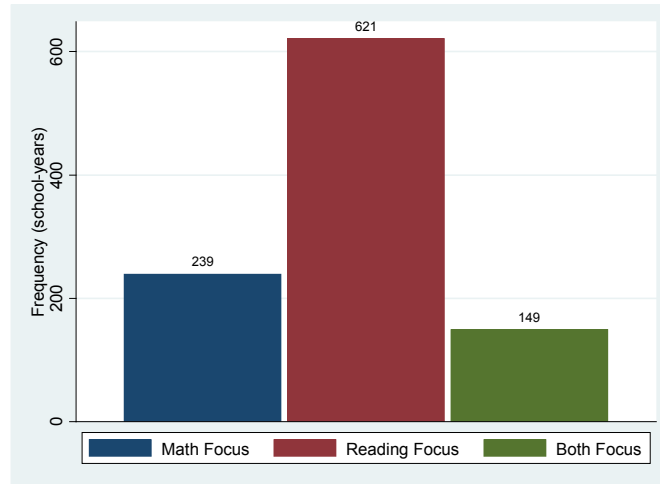


Figure 2.4: Test Focus Distribution (School*Years)

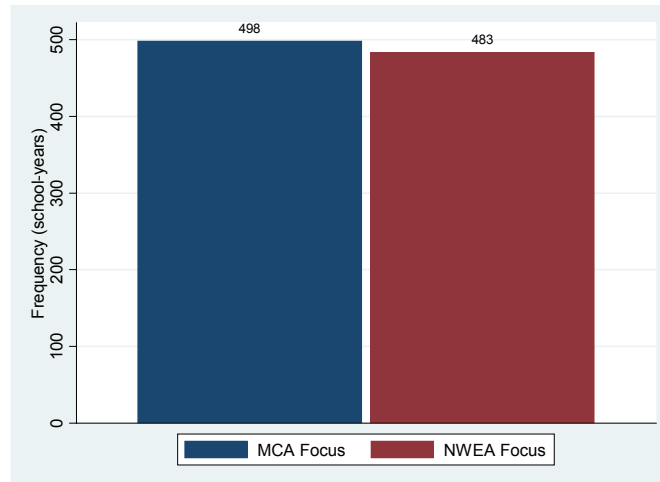


Figure 2.5: Program Type Distribution (Districts)

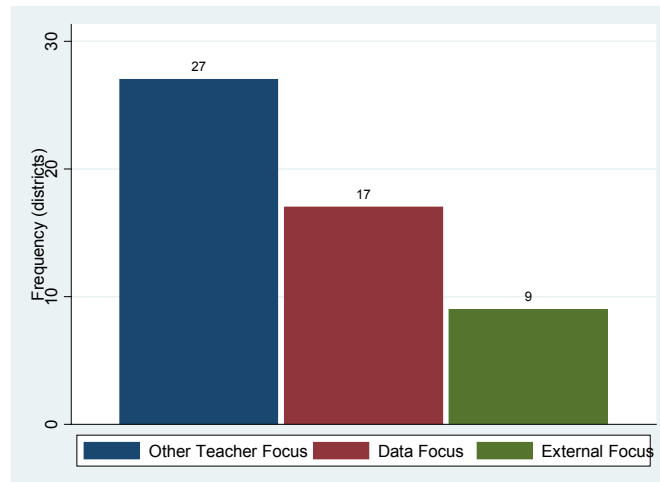


Table 2.1: Baseline Descriptives (Year=2005)		
	Mean (Standard Deviation)	Q-Comp Districts Non-Q-Comp Districts
District		
Student Enrollment	4791.019 (5538.579)	1810.444 (3953.147)
Student		
Percent Free/Reduced Lunch	16.898 (9.868)	23.255 (11.022)
Percent Special Education	11.637 (2.185)	12.856 (2.999)
Percent LEP	3.096 (4.860)	2.616 (5.255)
Percent Minority	11.927 (11.401)	9.765 (12.800)
Average MCA Math Score	1550.993 (44.585)	1529.481 (34.965)
Average MCA Reading Score	1573.984 (39.950)	1556.023 (35.121)
Teacher		
Average Years Experience	16.081 (2.738)	16.758 (2.570)
Percentage with Masters Degree	49.011 (17.780)	30.892 (17.471)
Average Teacher Salary	51353.545 (5750.179)	47016.710 (4890.432)

Table 2.2: Comparison of Full and Matched Sample		
	Mean (Standard Deviation)	MCA-NWEA Matched Sample
District		
Student Enrollment	2368.908 (4662.402)	2535.102 (4694.321)
Student		
Percent Free/Reduced Lunch	22.356 (11.218)	22.729 (11.124)
Percent Special Education	12.663 (2.912)	12.808 (3.112)
Percent LEP	2.757 (5.278)	2.591 (4.145)
Percent Minority	10.321 (13.009)	10.694 (11.412)
Average MCA Math Score	1532.747 (37.457)	1531.963 (36.306)
Average MCA Reading Score	1558.649 (36.835)	1555.616 (35.008)
Teacher		
Average Years Experience	16.635 (2.607)	16.648 (2.500)
Percentage with Masters Degree	33.851 (18.740)	34.885 (18.465)
Average Teacher Salary	47751.656 (5320.847)	47881.821 (5320.706)

Table 2.3: Q Comp Effects on Teacher Salary						
	Base Salary (1)	Bonus Salary (2)	Additional Pay (3)	Base Salary (4)	Bonus Salary (5)	Additional Pay (6)
Q Comp Effect	-2998.234** (1405.957)	755.646*** (141.474)	-265.555 (776.035)	-1228.604 (1403.370)	754.195*** (145.193)	-361.799 (763.113)
Teaching Experience Dummies				Y	Y	Y
Masters Degree Dummies				Y	Y	Y
Experience*Degree Interactions				Y	Y	Y
r ²	0.114	0.248	0.060	0.704	0.306	0.127
Q Comp Districts in Sample	15	15	15	15	15	15
Total Districts	63	63	63	63	63	63

All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Estimates are weighted at teacher level using SASS-provided weights. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.4: Q Comp Effects on Professional Development

Panel A: PD Hours	PD Hours -	PD Hours -	PD Hours -	PD Hours -
	Reading	Content Area	Technology	Classroom Management
	(1)	(2)	(3)	(4)
Q Comp Effect	3.120*	2.199	-1.424	-0.424
	(1.709)	(1.676)	(1.330)	(1.320)
r ²	0.164	0.132	0.130	0.098
Q Comp Districts in Sample	15	15	15	15
Total Districts	63	63	63	63

Panel B: PD Usefulness	Usefulness of Reading PD	Usefulness of Content PD	Usefulness of Tech PD	Usefulness of Mgmt PD
	(5)	(6)	(7)	(8)
Q Comp Effect	0.099	-0.009	0.075	0.168
	(0.141)	(0.115)	(0.214)	(0.207)
r ²	0.154	0.085	0.151	0.190
Q Comp Districts in Sample	15	15	15	15
Total Districts	63	63	63	63

Estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Estimates are weighted at teacher level using SASS-provided weights. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.5: Q Comp Effects on Teaching Climate					
	Administrative Support	Shared Values	Level of Satisfaction	Recognized for Work	Cooperative Effort
	(1)	(2)	(3)	(4)	(5)
Q Comp Effect	0.262**	0.070	0.178	0.352**	0.168
	(0.123)	(0.109)	(0.134)	(0.146)	(0.154)
r ²	0.157	0.128	0.114	0.138	0.127
Q Comp Districts in Sample	15	15	15	15	15
Total Districts	63	63	63	63	63

All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Estimates are weighted by teacher using SASS-provided weights. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.6: Q Comp Effects on Student Achievement

Panel A: MCA	MCA Math Scores			MCA Reading Scores			MCA Math Scores			MCA Reading Scores		
	Full Sample, Grades 3,5,7			Full Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q Comp Effect	0.025 (0.016)	0.021 (0.015)	0.024 (0.015)	0.010 (0.011)	0.008 (0.011)	0.011 (0.011)	0.013 (0.020)	0.011 (0.019)	0.029* (0.016)	0.013 (0.012)	0.011 (0.013)	0.023 (0.018)
Q_2ormore dummy		-0.043** (0.017)			-0.022 (0.017)			-0.034* (0.019)			-0.027 (0.027)	
Pretrend included			Y			Y			Y			Y
r2	0.834	0.835	0.835	0.862	0.862	0.862	0.832	0.833	0.790	0.863	0.864	0.802
student-years	366763	366763	366763	373432	373432	373432	238747	238747	154606	244609	244609	156845
N (school-grade-years)	3277	3277	3277	3278	3278	3278	2036	2036	1306	2036	2036	1308
Panel B: NWEA	NWEA Math Scores			NWEA Reading Scores			NWEA Math Scores			NWEA Reading Scores		
	NWEA Sample, Grades 3-8			NWEA Sample, Grades 3-8			Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q Comp Effect	0.030 (0.035)	0.029 (0.034)	0.049 (0.029)	-0.003 (0.016)	-0.003 (0.016)	0.002 (0.013)	0.054 (0.062)	0.052 (0.056)	0.085* (0.048)	0.008 (0.027)	0.014 (0.028)	0.007 (0.023)
Q_2ormore dummy		-0.019 (0.030)			0.015 (0.017)			-0.012 (0.042)			0.032 (0.023)	
Pretrend included			Y			Y			Y			Y
r2	0.918	0.918	0.910	0.901	0.901	0.891	0.946	0.946	0.938	0.937	0.937	0.928
N (student-years)	316032	316032	219819	316379	316379	219588	198501	198501	138764	198632	198632	138269

All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.7: Q Comp Effects across Achievement Distribution

Panel A: MCA		<50% Proficient		>85% Proficient	
		Full Sample, Grades 3,5,7		Full Sample, Grades 3,5,7	
<i>Math Effect</i>		0.143***	0.143***	-0.031	-0.039
		(0.049)	(0.048)	(0.044)	(0.040)
Q_2ormore dummy			-0.001		-0.091***
			(0.037)		(0.021)
N (school-grade-years)		504	504		
<i>Reading Effect</i>		0.125*	0.118*	-0.032**	-0.035**
		(0.065)	(0.059)	(0.014)	(0.015)
Q_2ormore dummy			-0.060		-0.031
			(0.050)		(0.027)
N (school-grade-years)		217	217	1170	1170

Panel B: NWEA		Lowest Quartile		Second Lowest Quartile		Second Highest Quartile		Highest Quartile	
		NWEA Sample, Grades 3-8		NWEA Sample, Grades 3-8		NWEA Sample, Grades 3-8		NWEA Sample, Grades 3-8	
<i>Math Effect</i>		0.293***	0.238***	0.072***	0.063**	-0.043	-0.026	-0.196***	-0.141***
		(0.026)	(0.032)	(0.024)	(0.028)	(0.033)	(0.036)	(0.039)	(0.036)
Q_2ormore dummy			0.227***		0.039		-0.063**		-0.229***
			(0.033)		(0.029)		(0.030)		(0.036)
N (student-years)		27425	27425	28788	28788	29127	29127	29434	29434
<i>Reading Effect</i>		0.342***	0.272***	0.078***	0.064***	-0.073***	-0.056***	-0.274***	-0.211***
		(0.023)	(0.021)	(0.015)	(0.013)	(0.015)	(0.015)	(0.017)	(0.013)
Q_2ormore dummy			0.287***		0.054**		-0.067***		-0.248***
			(0.021)		(0.024)		(0.022)		(0.018)
N (student-years)		27696	27696	28900	28900	29401	29401	29286	29286

All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.8: Q Comp Effects across Student Subgroups

Panel A: MCA	Male		Female		White	
	Full Sample, Grades 3,5,7		Full Sample, Grades 3,5,7		Full Sample, Grades 3,5,7	
<i>Math Effect</i>	0.028*	0.024*	0.022	0.019	0.025	0.021
	(0.014)	(0.014)	(0.018)	(0.018)	(0.016)	(0.016)
Q_2ormore dummy		-0.045***		-0.037		-0.039**
		(0.014)		(0.022)		(0.015)
N (school-grade-years)	3457	3457	3439	3439	3343	3343
<i>Reading Effect</i>	0.009	0.007	0.022	0.020	0.015	0.014
	(0.011)	(0.012)	(0.014)	(0.015)	(0.010)	(0.010)
Q_2ormore dummy		-0.026		-0.021		-0.013
		(0.021)		(0.014)		(0.015)
N (school-grade-years)	3472	3472	3453	3453	3342	3342
Panel A: MCA (continued)	Non-White		Free/Reduced Lunch			
	Full Sample, Grades 3,5,7		Full Sample, Grades 3,5,7			
<i>Math Effect</i>	0.026*	0.023	0.019	0.014		
	(0.014)	(0.014)	(0.022)	(0.023)		
Q_2ormore dummy		-0.040**		-0.050**		
		(0.016)		(0.025)		
N (school-grade-years)	5677	5677	2791	2791		
<i>Reading Effect</i>	0.014	0.012	0.020	0.012		
	(0.011)	(0.012)	(0.025)	(0.025)		
Q_2ormore dummy		-0.028*		-0.077***		
		(0.016)		(0.023)		
N (school-grade-years)	5976	5976	2841	2841		

Continued on next page

Table 2.8 (Continued): Q Comp Effects across Student Subgroups								
Panel B: NWEA	Male		Female		White		Non-White	
	NWEA Sample, Grades 3-8		NWEA Sample, Grades 3-8		NWEA Sample, Grades 3-8		NWEA Sample, Grades 3-8	
<i>Math Effect</i>	0.024	0.023	0.036	0.035	0.025	0.025	0.011	0.011
	(0.036)	(0.034)	(0.035)	(0.034)	(0.034)	(0.033)	(0.042)	(0.041)
Q_2ormore dummy		-0.019		-0.020		-0.019		-0.009
		(0.032)		(0.028)		(0.032)		(0.032)
N (student-years)	161406	161406	154626	154626	246904	246904	56840	56840
<i>Reading Effect</i>	-0.008	-0.007	0.001	0.001	-0.010	-0.010	-0.004	-0.004
	(0.019)	(0.020)	(0.014)	(0.014)	(0.014)	(0.014)	(0.018)	(0.019)
Q_2ormore dummy		0.024		0.006		0.015		0.021
		(0.021)		(0.016)		(0.016)		(0.029)
N (student-years)	161636	161636	154743	154743	246684	246684	57465	57465

Samples for each regression are limited to students in the subgroup. All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.9: Q Comp Effects by Program's Subject Focus

Panel A: MCA	MCA Math Scores			MCA Reading Scores			MCA Math Scores			MCA Reading Scores		
	Full Sample, Grades 3,5,7			Full Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q*Math Focus	0.009 (0.024)	0.005 (0.022)	0.010 (0.024)	0.000 (0.018)	-0.001 (0.018)	0.002 (0.018)	0.013 (0.028)	0.012 (0.027)	0.013 (0.028)	-0.002 (0.019)	-0.004 (0.019)	-0.002 (0.018)
Q*Reading Focus	-0.011 (0.018)	-0.013 (0.017)	-0.008 (0.018)	0.023 (0.015)	0.022 (0.015)	0.026* (0.015)	-0.008 (0.024)	-0.010 (0.024)	-0.008 (0.024)	0.020 (0.017)	0.019 (0.017)	0.020 (0.017)
Q*Both Focus	0.004 (0.020)	0.002 (0.019)	0.007 (0.023)	0.019 (0.015)	0.018 (0.015)	0.023 (0.015)	0.003 (0.027)	0.003 (0.026)	0.003 (0.027)	0.015 (0.016)	0.015 (0.016)	0.016 (0.015)
Q_2ormore dummy	Y			Y			Y			Y		
Pretrend Included	Y			Y			Y			Y		
r2	0.868	0.868	0.869	0.890	0.890	0.890	0.872	0.872	0.872	0.895	0.895	0.895
N (school-grade-years)	3034	3034	3034	3034	3034	3034	1901	1901	1901	1897	1897	1897

Panel B: NWEA	NWEA Math Scores			NWEA Reading Scores			NWEA Math Scores			NWEA Reading Scores		
	NWEA Sample, Grades 3-8			NWEA Sample, Grades 3-8			Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q*Math Focus	0.006 (0.030)	0.006 (0.030)	0.014 (0.022)	-0.015 (0.018)	-0.015 (0.018)	-0.017 (0.024)	-0.010 (0.063)	-0.011 (0.059)	-0.009 (0.046)	-0.020 (0.029)	-0.015 (0.028)	-0.033 (0.029)
Q*Reading Focus	0.026 (0.033)	0.026 (0.032)	0.023 (0.034)	0.002 (0.022)	0.002 (0.021)	0.000 (0.025)	0.030 (0.066)	0.029 (0.062)	0.015 (0.056)	0.003 (0.039)	0.008 (0.037)	-0.006 (0.034)
Q*Both Focus	-0.014 (0.040)	-0.014 (0.040)	-0.020 (0.040)	-0.008 (0.023)	-0.008 (0.023)	-0.013 (0.025)	-0.016 (0.063)	-0.018 (0.061)	-0.033 (0.062)	-0.018 (0.034)	-0.010 (0.033)	-0.026 (0.035)
Q_2ormore dummy	Y			Y			Y			Y		
Pretrend included	Y			Y			Y			Y		
r2	0.916	0.916	0.910	0.898	0.898	0.892	0.943	0.943	0.938	0.935	0.935	0.929
N (student-years)	285775	285775	201772	286284	286284	201316	179905	179905	127954	180006	180006	127298

All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.10: Q Comp Effects by Program's Test Focus

Panel A: MCA	MCA Math Scores			MCA Reading Scores		
	Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q*MCA Focus	-0.001 (0.026)	-0.002 (0.026)	-0.001 (0.026)	0.020 (0.013)	0.018 (0.014)	0.020 (0.013)
Q*NWEA Focus	0.001 (0.023)	0.001 (0.022)	0.001 (0.022)	0.006 (0.017)	0.005 (0.017)	0.006 (0.017)
Q_2ormore dummy		Y			Y	
Pretrend Included			Y			Y
r2	0.871	0.872	0.871	0.894	0.895	0.895
N	1901	1901	1901	1897	1897	1897

Panel B: NWEA	NWEA Math Scores			NWEA Reading Scores		
	Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q*MCA Focus	-0.005 (0.091)	-0.006 (0.093)	-0.063 (0.068)	-0.009 (0.040)	-0.001 (0.040)	-0.011 (0.035)
Q*NWEA Focus	-0.014 (0.073)	-0.015 (0.071)	-0.057 (0.054)	-0.008 (0.030)	-0.003 (0.028)	-0.023 (0.026)
Q_2ormore dummy		Y			Y	
Pretrend included			Y			Y
r2	0.943	0.943	0.938	0.935	0.935	0.929
N	179905	179905	127954	180006	180006	127298

All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

Table 2.11: Summary Statistics by Program Type (Year=2005)

Mean (Standard Deviation)	Other Teacher Focus	Data Focus	External Focus
District			
Student Enrollment	5510.407 (6953.293)	3945.353 (3951.565)	5576.444 (3769.628)
Student			
Percent Free/Reduced Lunch	17.745 (10.757)	13.993 (8.965)	18.064 (6.143)
Percent Special Education	11.419 (1.953)	11.769 (2.550)	11.338 (1.146)
Percent LEP	3.786 (6.173)	1.573 (2.075)	4.759 (3.978)
Percent Minority	13.351 (13.687)	8.575 (7.468)	15.414 (10.892)
Average MCA Math Score	1551.866 (50.693)	1562.880 (35.626)	1531.704 (25.329)
Average MCA Reading Score	1573.307 (46.508)	1585.003 (30.629)	1553.487 (25.178)
Teacher			
Average Years Experience	16.130 (2.727)	16.536 (3.098)	15.196 (2.586)
Percentage with Masters Degree	48.568 (19.808)	49.329 (16.746)	56.119 (9.403)
Average Teacher Salary	50907.232 (6567.791)	51259.480 (5033.232)	53635.672 (4760.223)

Table 2.12: Q Comp Effects by Program Type

Panel A: MCA	MCA Math Scores			MCA Reading Scores			MCA Math Scores			MCA Reading Scores		
	Full Sample, Grades 3,5,7			Full Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q*Other Teachers	0.016 (0.014)	0.010 (0.014)	0.011 (0.014)	0.002 (0.013)	-0.001 (0.014)	0.007 (0.013)	0.019 (0.028)	0.015 (0.028)	0.041 (0.029)	Matched Sample, Grades 3,5,7		
Q*Data	0.035 (0.031)	0.034 (0.032)	0.037 (0.031)	0.025 (0.018)	0.024 (0.018)	0.032* (0.017)	-0.002 (0.030)	-0.003 (0.030)	0.019 (0.027)	0.004 (0.018)	0.004 (0.019)	0.018 (0.030)
Q*External	0.057*** (0.018)	0.059*** (0.018)	0.059*** (0.018)	-0.002 (0.019)	-0.002 (0.019)	-0.013 (0.019)	0.020 (0.022)	0.024 (0.024)	0.018 (0.027)	0.006 (0.030)	0.009 (0.027)	-0.002 (0.039)
Q_2ormore dummy	Y			Y			Y			Y		
Pretrend included	Y			Y			Y			Y		
r2	0.869	0.869	0.869	0.891	0.891	0.891	0.872	0.872	0.791	0.895	0.895	0.804
N (school-grade-years)	3086	3086	3086	3086	3086	3086	1907	1907	1177	1906	1906	1178

Panel B: NWEA	NWEA Math Scores			NWEA Reading Scores			NWEA Math Scores			NWEA Reading Scores		
	NWEA Sample, Grades 3-8			NWEA Sample, Grades 3-8			Matched Sample, Grades 3,5,7			Matched Sample, Grades 3,5,7		
Q*Other Teachers	-0.004 (0.035)	-0.007 (0.033)	0.013 (0.027)	-0.006 (0.022)	-0.005 (0.022)	0.002 (0.020)	0.021 (0.065)	0.015 (0.062)	0.047 (0.051)	0.014 (0.040)	0.022 (0.042)	0.019 (0.043)
Q*Data	0.098* (0.053)	0.098* (0.052)	0.115** (0.050)	0.030 (0.026)	0.030 (0.025)	0.027 (0.027)	0.126 (0.090)	0.123 (0.085)	0.158** (0.073)	0.035 (0.039)	0.039 (0.036)	0.035 (0.037)
Q*External	0.006 (0.031)	0.011 (0.034)	0.005 (0.031)	-0.011 (0.036)	-0.014 (0.034)	-0.001 (0.043)	-0.006 (0.045)	-0.003 (0.051)	-0.008 (0.042)	-0.018 (0.063)	-0.022 (0.056)	-0.025 (0.081)
Q_2ormore dummy	Y			Y			Y			Y		
Pretrend included	Y			Y			Y			Y		
r2	0.917	0.917	0.910	0.899	0.899	0.891	0.945	0.945	0.938	0.936	0.936	0.928
N (student-years)	300329	300329	209203	300674	300674	208944	188888	188888	132376	188972	188972	131848

All estimates control for student population characteristics including race, gender, percent free/reduced-price meals, percent special education, and percent LEP. Robust standard errors included in parentheses are clustered at the district level. *** p<0.01 ** p<0.05 * p<0.1.

CHAPTER THREE

ALIGNING TEACHER IMPROVEMENT STRATEGIES AT THE STATE LEVEL: THE DESIGN AND IMPLEMENTATION OF Q COMP IN MINNESOTA

I. INTRODUCTION

One of the most ambitious initiatives to emerge from the policy pipeline in recent years is the teacher improvement system (TIS). By TIS, I refer to a type of program that attempts to systemically transform teachers' incentives, their learning opportunities, and the feedback they receive. Current examples include IMPACT in Washington D.C., Pro-Comp in Denver, Colorado, and the Teacher Advancement Program, which operates in over 50 districts across 13 states.

The potential of such programs derives from the attention they give to teaching itself as the major lever of school improvement. Although repeated studies have demonstrated that instructional quality has a major impact that on student achievement, teachers still tend to operate within a system that offers little support for their work (Cohen, 2011; Elmore, 2004; Lortie, 1975). In contrast, TISs offer a compelling theory of change. They act upon the simple but powerful theory that improvement in teaching depends not on any single factor but on the interaction between teachers' incentives and their opportunities to learn on the job. Thus, TISs attempt to create a coherent structure around teacher growth and development, aligning the process so that teachers work within a professional culture of continuous improvement.

These programs are now rapidly transitioning from the district to the state level as a result of several parallel forces including federal pressure through initiatives like the Race to the Top

grant competition, sponsorship from a number of influential non-profit foundations, and support for such policies from both major political parties. States that have pledged to implement systems that integrate teacher evaluation, professional development, and salary reform in the next several years include Tennessee, Delaware, Wisconsin, Connecticut, and several others.

Yet as these states begin the process of program design, they face substantial challenges. First, there has been little demonstration of success for these efforts at any level of governance so far (Glazerman & Seifullah, 2010; Podgursky & Springer, 2007). Equally importantly, state central offices have not had great success running programs that aim to reform the organization and work of teaching, and there has been relatively little research as to what is feasible and what is likely for such programs when directed from this vantage point.

Historically, most states have delegated most control over on-the-job teacher policy to districts, which have either passed them on to school-level administrators or dictated them through collective bargaining with local unions. For agreements about teacher career progressions or professional development to be even partially negotiated at the state level represents a major break with the past and a potential source of considerable political controversy. As the policy meets the political, there is a real question about what these teacher improvement systems will become.

My intent here is to examine the ways that the intersection between governance and policy might play out across the nation and what it might mean for the likelihood that TISs will effect genuine change that reaches beyond administration and into actual teacher performance.

To frame the work, I consider the theory behind teacher improvement policies within the context of what we know about the interplay between state and local policymaking in the U.S. In the body of the article, I make use of the fact that the state of Minnesota took a turn toward such

policies in 2005, more than half a decade before the broader trend in this direction. Using Minnesota as a case study and illustrative example, I am able to look at the way several of the major dilemmas around TISs are playing out in policy and practice and to highlight both some of the predictable and more surprising outcomes to emerge from the process.

The example offers some suggestion of the challenges and opportunities associated with recent state-based human capital strategies. I suggest that the success of these programs depends in large part on whether they transmit new instructional expertise to the teachers in participating districts, but that this transmission is not a natural consequence of the TIS structure. Instead, it depends on levels of support for teacher learning that can be provided by the various organizations involved. This analysis raises serious questions about the degree to which these new reform policies will bring about the desired changes in teacher practice or student learning.

II. THEORETICAL FRAME

As noted in the introduction, I use the term teacher improvement system to refer to a specific type of program that draws several human capital reform strategies into a single, interconnected structure. Broadly, such programs tend to be characterized by three major parts. The first is a salary and/or tenure schedule that attempts to increase teachers' accountability to some form of performance evaluation. The second is a teacher career ladder that rewards teacher expertise and performance by allowing classroom teachers to take on extra responsibilities including peer observation, mentoring, and/or learning team leadership. The third is a professional development process tied both to the performance goals of the first piece and the observation and learning teams of the second piece. While the individual elements within these systems are not new, their combination, in coordination with the school test accountability

ushered in by No Child Left Behind, represents a recent development in the longstanding search for policies that can influence teacher performance.

The spread of these programs and the move to implement at the state level owes much to a concerted federal effort. Through the grant competition known as the Race to the Top, states have been asked to submit plans that include an aligned approach to teacher evaluation, promotion, and professional development. Similarly, the federal Teacher Incentive Grant program, which initially pushed for simple pay-for-performance reforms, has been amended to call for a more systemic approach that includes many of the same types of initiatives. Even outside Washington, the trend in this direction has been apparent for some time. Over the past decade, a series of private foundations have taken their own steps to transform teacher support and development. Groups like the Teacher Advancement Program, The New Teacher Project and the Bill and Melinda Gates Foundation have created initiatives to promote the growth of TISs in their many partner districts. And, as noted earlier, several districts like Denver, CO and Washington, D.C. have independently designed their own such programs.

Though different in detail, these reforms share a common base and theoretical perspective. In essence, the programs attempt to drive teacher and student improvement by drawing tighter connections between students' achievement, teachers' rewards and recognition, and the teacher learning that makes increased student achievement possible (see Figure 3.1).

To define these reforms as systems is to specify a genre of reform that up until now has been defined mostly by its individual parts. Many such programs are characterized as merit pay or pay-for-performance programs, even when pay reform might only be a relatively small piece of the intervention (Sawchuk, 2009). At other times, the teacher support elements are sometimes picked out as the single component that defines the reform. However, simplifying these

programs into single parts risks overlooking their greatest assets, namely the connections across individual elements. Indeed, the crucial argument made by many who espouse these new reforms is that they have the potential to be more than the sum of their parts.

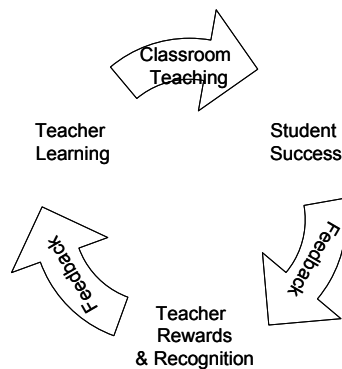


Figure 3.1: Teacher Improvement System Schematic

A set of concepts for elucidating this view comes from the field of business economics, where researchers have tried to capture the complex relationship between incentives, effort, learning, and output (Siemson, Roth, & Sridhar, 2008; Bonner & Sprinkle, 2002; Camerer & Hogarth, 1999). According to this literature, incentives – whether in the form of concrete rewards or broader environmental conditions – are only useful under certain conditions. They help to clarify the goals of a particular task, but when applied in isolation they produce significant positive results only on tasks that demand little specialized knowledge. For more specialized tasks, another variable comes into play, specifically the “cognitive capital” that a subject either possesses or has the opportunity to acquire. Cognitive capital is used here to represent knowledge that has been tailored to particular situations, the “variety of tricks or approaches to solving an experimental task, like the many specialized tools on a carpenter’s tool belt or a cook’s knowledge of foods, kitchen utensils, and recipes” (Camerer & Hogarth, 1999, p.

9). The term is particularly useful for describing the elusive combination of knowledge and know-how that characterizes teaching expertise, a combination that is not easily captured by typical measures of human capital (Shulman, 1987). For specialized tasks that require high degrees of cognitive capital, incentive effects tend to be far less predictable, producing non-existent or even negative outcomes, unless they are accompanied by opportunities for subjects to enhance their abilities to meet the incentivized goals.

Applying the theory of cognitive capital to teaching and teacher improvement provides a useful way of conceptualizing teaching improvement policy. Teaching reform is often couched in terms of competing theories, with one line of thought proposing the need for stronger incentives tied to student achievement and another suggesting that teachers need better guidance and professional development in order to succeed. Yet it seems far more likely that the two strategies are complements not substitutes. If teachers lack the capabilities or cognitive capital to substantially improve their practice – and there is substantial evidence that they do – then even powerful incentives will appear ineffective, just as professional development opportunities cannot ensure better instruction if many teachers lack the incentives to apply concerted effort to this task.

At its core, the promise of a TIS is that it can bridge this gap, creating a more coherent and aligned structure for teacher on-the-job development where teachers agree upon a common set of goals and find genuine opportunities to build their knowledge of how to meet these goals. Such a goal seems particularly necessary and important in the realm of K-12 education where policy analysts have repeatedly pointed out the contradictions in the way human capital policies have been traditionally defined (Odden, 2008). Yet despite its apparent logic, the task seems

daunting when undertaken within our current educational structure, especially from the state level where many of the recent initiatives are based.

Researchers who have studied the vast catalogue of large-scale programs that have tried and failed to spur genuine classroom improvement often explain these outcomes as a consequence of the uniquely structured system of U.S. educational governance (Cohen & Moffitt, 2009; Elmore, 1996). While school control is nominally centered at the state level, funding and decision-making gets spread across multiple agencies and layers of governance. Early implementation research by RAND scholars into a series of federally sponsored initiatives noted that the degree of local autonomy in the educational system generated an adaptive process that often rendered initial program goals unrecognizable (Berman & McLaughlin, 1977; McLaughlin, 1987). Other researchers have documented the ways that understandings of policy directives change as they filter through different agencies and subunits (Cohen & Hill, 2001; Spillane, 2000; Spillane & Jennings, 1997). Although standards-based reform has offered some uniformity at the level of testing and accountability, many of the decisions pertaining to teacher quality, including curriculum selection, professional development plans, and teacher assignment get made at the district, school, and classroom levels (Cohen & Spillane, 1993).

A central consequence of this structure is that it becomes difficult for any single agency such as a state department to find purchase around its instructional guidance. Cohen (2011) describes this as a problem of infrastructure. States and the federal government lack the supporting arrangements that would enable strong teacher support. A concrete example can be found in the realm of professional development. Since curriculum and teaching frameworks are not necessarily shared across districts or even schools, guidelines must remain vague and large-scale professional development cannot reference particular texts or student tasks. Divided

funding sources contribute to multiple, and often contradictory goals for professional development within districts (Miles et al., 2004). Similarly, expertise around particular teaching techniques is diffused throughout the system, often in ways that reflect enduring inequalities between schools and districts (Lankford, Loeb, & Wyckoff, 2002). In this context, state departments of education, far from serving as centralized repositories of expertise, have traditionally remained quite weak with respect to their instructional capability. Compounding the problem, they have been provided with little in the way of staff or funding to offer more robust forms of guidance and support.

From one perspective, TISs have the potential to solve some of these issues. By creating a structure that links teacher performance to salary, leadership, and professional development, TISs open the door to new kinds of conversations around teaching improvement that span classrooms and schools. Yet these new structures are also likely to directly collide with the fragmentation of the greater system. The vision of coherence contained within the TIS structure demands greater alignment between teachers' incentives and their opportunities to acquire cognitive capital around particular teaching goals. But this in turn demands initial consensus around ideas about teacher performance (e.g. what should be taught and how it should be measured) and about teacher learning (e.g. what types of learning are valuable and how these can be transmitted), two areas where consensus has been difficult to obtain. How states obtain this consensus and the extent to which the process changes the tenor of instructional guidance represent the central issues in TIS design and implementation.

To understand more about the practical implications of such concerns, I turn to one of the few fully developed examples of a state-directed TIS that currently exists. As noted in the introduction, Minnesota adopted Q Comp in 2005, creating a program that closely resembles the

TISs that are to be adopted more widely over the next few years. In doing so, state officials had to confront the issues outlined above. The eventual compromises offer a study of what such reforms actually look like when created and operated within our complex system of educational governance, as well as a possible illustration of the expected and unexpected ways that the move toward teacher improvement systems is likely to play out as these initiatives spread across the country.

III. STUDY AND METHODOLOGY

My investigation of Minnesota's teacher reform policies adopts a case study approach designed to probe individuals' understandings of a particular set of circumstances (Merriam, 1997). Drawing on both interviews and documentary sources, the research includes a historical component concerning the passage of the Q Comp legislation and its initial implementation and a current component concerning the day-to-day operation of the program.

Documentary sources relating to the history of the program include bills from the Minnesota legislature and summaries of legislative proceedings, audiotaped transcripts of legislative hearings, organizational statements and news releases, and news articles generated by a Lexis-Nexis search. These documents helped me to put together an initial timeline of events and to trace the official positions of relevant organizations. They also helped me identify key figures and the roles they played across various stages of Q Comp's existence.

These names served as my initial set of interview contacts as I worked to gather more detail about the ways that events unfolded and the reasons for this progression. From here, I used a snowball technique to identify other figures who were involved either directly or behind the scenes in defining Q Comp's course of development. This process generated a set of primary

subjects, several of whom were interviewed repeatedly over the course of the project. Key subjects included the primary authors of the program legislation in the State House and Senate, members of the former Governor's office, the current and former central leadership of the Minnesota Department of Education (MDE), and relevant officials from Education Minnesota, the state teachers union. A full list of interview subjects is provided in Appendix A. Interviews were semi-structured and covered topics that included the program's goals, the legislative process, and the major influences on program policies.

When relevant, I also asked interview subjects about the current workings of the program. In trying to understand more about program operations, I benefitted from an extensive set of programmatic materials made available by MDE, including all current and former district applications, approval letters, goal updates, and plan revisions, as well as FAQs, sample rubrics, and guidelines for Q Comp districts. Similarly, officials at Education Minnesota provided several of the training materials that the organization had used when working with districts about Q Comp.

I adopted an integrated process of data collection and analysis, allowing me to develop working hypotheses that I then tested as my circle of interview subjects grew (Miles & Huberman, 1994). I analyzed documents and transcripts using a process of open and iterative coding to draw descriptive themes from the data (Corbin & Strauss 2008). These themes generated axial codes that allowed me to group responses into several major categories—state vs. local control; power shifts; MDE enforcement; MDE capacity; funding concerns, and; collaboration—and selectively code within these categories. Using the coded data, I was able to draw a rich picture of the ways that different constituents viewed, interpreted, and acted upon the events surrounding Q Comp.

My work here is additionally informed by my ongoing broader study of teacher policy reform in Minnesota that includes an evaluative component looking at the Q Comp's effects on student test scores and case studies of three Q Comp districts. The evaluative component provided information about the limited causal impact of the Q Comp policy, and the case studies offered useful detail about how state actions have been interpreted and implemented at the district level.

IV. FINDINGS

Minnesota's struggles to reform teacher human capital policies within the structure of a TIS are likely to be duplicated across RTTT states. Bipartisan compromise produced legislation that defined a basic program structure, which included teacher observation and evaluation, performance pay, and teacher-led professional development components and requirements for aligning these parts in pursuit of well-specified improvement goals, but much remained unspecified in legislation. The responsibility for the details of these arrangements fell both to the Department of Education and to the individual districts that submitted program applications.

The subsequent division of this responsibility both bears out the relationships that I reviewed in my theoretical framework and provides further insight into the workings of a state-based TIS. At the beginning of the program, much of the department's energy was devoted to political wrangling around district salary schedules. Motivated by the promise of state-provided funding for teacher salaries, most districts hoped to spread the money to as many teachers as possible while the state pushed for more drastic versions of merit pay. The compromises that eventually emerged created a standardized set of understandings around pay reform but resulted in pay structures within most districts that minimized the amount of teacher pay based directly

on student performance. This outcome follows a common trend where differential compensation plans tend to be used as the means for supporting particular types of teacher work and for creating teacher buy-in rather than the outcome-based rewards that policymakers tend to envision (Murnane & Cohen, 1986).

Under these circumstances, Q Comp's evaluation and professional development structures took on a central role, since districts' pay reforms primarily supported teacher participation in these elements of the program. In many ways, officials in the department of education took an aggressive stance toward program standardization, defining structural requirements around teacher observations, learning team meetings, and goal-setting procedures that clarified much of the ambiguity in the original legislation. Moreover, notwithstanding staff and funding constraints in MDE, these officials worked extensively with individual districts on applications and yearly goal-setting procedures to ensure that the programs fit these requirements and demonstrated alignment with school and district goals.

However the structural requirements from MDE did not provide a source of expertise about teaching, and MDE offered little further guidance in this department. As was to be expected within a system that includes few tools for centralized instructional support, MDE left Q Comp teacher-leaders, union representatives, and district administration to determine how to make professional development relevant, informative, and useful. Interestingly, the teachers union appeared poised at times to serve as an important supplementary source of instructional expertise, filling in where MDE could not. Yet the relationship between the state union and MDE did not last, leaving much of the responsibility for programmatic success or failure at the district level in the hands of teachers, local union representatives, and district administrators. This outcome points to the major hurdle in TIS reform at the state level. Although the TIS provides a

useful structure and extra funding for teacher learning to take place, the structure does not in itself solve the problem of instructional guidance. In Minnesota, this problem was left to districts to solve, just as it had been before the program began.

In the sections below, I develop this narrative in greater detail. After briefly summarizing the context of the Minnesota Q Comp program and what is known about the program's impact on student achievement, I describe the actual intervention, in legislation and as it was taken up by MDE and the teachers union. Finally, I investigate the implications of this analysis. In particular, my concluding sections begin to explore the types of policy and research decisions that might promote greater instructional learning within such programs.

A. Program Context and Results

By the turn of the millennium, Minnesota possessed a number of unique characteristics that help to explain its early shift in the direction of teacher pay and professional development reform. The state took great pride in its strong legacy of educational support, and officials at all levels of government tended to reference public education when describing the state's core values. Through the late 1990s and early 2000s, Minnesota legislators hosted a steady procession of policy experts who offered prescriptions for keeping Minnesota at the forefront of educational reform. Many of these experts, including Allen Odden from the Center for Policy Research in Education (CPRE), Lowell Milken, the founder of the Milken Family Foundation and the TAP program, and Brad Jupp who worked for the Denver Public Schools before joining the federal government, held a particular interest in reforming the "steps and lanes" teacher pay scale that could be found in nearly every Minnesota district.

At the same time as consensus grew around the need for school reform, the dominant atmosphere was one of growing partisanship caused by rapid shifts in political power at each level of government. The governorship shifted from the reform party's Jesse Ventura to the Republican Tim Pawlenty, the House oscillated between parties, and the Senate remained in the hands of a strong Democratic-Farmer-Labor majority. State funding for schools became a crucial area of dispute as a series of budget crises prompted repeated cuts to previously secure funding streams.

To the conservatives in the Governor's office, "merit pay" reforms that would tie teacher compensation to classroom test results seemed a clear political winner on the state and national stages, and both Ventura and then Pawlenty spoke favorably of such legislation. Yet strict pay-for-performance standards seemed unlikely to get through the Democratic Senate, which tended to look skeptically on proposals that relied too heavily on test results. Moreover, Minnesota possessed the distinction of being one of the few states in the country where the two major teachers unions—the National Education Association and the American Federation of Teachers—had successfully merged into a united organization. The powerful state union had early on signaled its own intent to be engaged in any reform effort by publishing a list of principles that would need to be included within any alternative compensation plan that the organization would support (Education Minnesota, 2000). These principles demanded strong local control and professional development support, making many merit pay reforms unlikely.

Education Minnesota's principles helped keep strict merit pay off the bargaining table, but by setting out favored policies, the document also helped clear the way for other legislation. Lawmakers now had room to craft proposals that they did not have to fear would provoke immediate opposition. As school budget concerns continued to grow, the idea of some form of

teacher pay reform coupled with increased state support for teacher and district development offered a potential compromise that Democrats and Republicans could agree upon to break the impasse over state school funding.

Starting in 2001, the legislature agreed to fund a low-stakes experiment to spread several million dollars between a handful of districts that hoped to experiment with alternative pay and professional development schemes. This experiment contributed to a general feeling over the next few years that such reform was on its way, and that the best move for those who questioned some of its tenets was not direct opposition but rather a concerted push to shape the next stage in ways that would be acceptable to all parties. Thus, in 2005 in a year of new budget negotiations, a growing consensus between legislative parties, the governor's office, and state union representatives set the stage for further action. In the House of Representatives, Republican leaders drafted their version of a teacher pay reform bill in collaboration with the Governor's office while D-F-L senators filed their own take on the policy. When budget negotiations pushed the legislature into special session, the bills came together in July 2005 to create the program named by law as the Alternative Teacher Professional Pay System and renamed by the Minnesota Department of Education (MDE) as Quality Compensation for Teachers or Q Comp.

Seven years after the passage of this legislation, views on Q Comp remain divided. On the one hand, the program has been something of a disappointment on a global level. When the state initially appropriated \$90 million for the program with a specified per-pupil grant for each joining district, legislators were concerned that a run on the funds could leave some districts empty-handed. Instead, the program never reached capacity. Growth across traditional (non-charter) districts has proceeded quite slowly, with only 47 of the state's 364 traditional school

districts in the program by 2010 (see Figure 3.2).¹ Overall, the districts that have joined the program tend to be more affluent, more urban, and more populated than those that did not.

Equally importantly, the program that has drawn on over \$200 million in state funds during its tenure has shown little evidence of large-scale improvements in student achievement. Two recent evaluations of state tests suggest only small average test score gains of in adopting districts as compared with those that did not join the program (Sojourner, West, & Mykerezi, 2012; Schwartz, 2012).²

However, statewide evaluations of the program hide some of the more local successes. While there have been no large gains in achievement on average, there is considerable variation in improvement across districts suggesting the possibility that some districts have benefitted considerably from the program while others have not—an outcome that might have been expected from a program with substantial local autonomy. There is also evidence that teachers and administrators within adopting districts have found much to like in the program (Hezel Associates, 2009), and nearly all of adopters have remained in the program to date. Even focusing on the low number of adopting districts in the state misses some detail. Because the districts that have adopted the program have tended to be larger than average, the 45 districts that have joined encompass more than one third of the state's students, giving the program a substantial reach in a large and educationally influential state.

In this context, many policymakers within the state tend to view the program not as a final answer to the problem of teacher improvement but rather as a useful step in the direction of

¹ Charter schools also have the option of joining Q Comp, and 27 joined between 2006 and 2010. Since charters have fewer initial constraints around teacher pay and support and therefore face a lower opportunity cost to join the program, my analysis focuses on traditional districts.

² Sojourner, West, & Mykerezi (2012) argue that the student gains they identify (~0.03 standard deviations in reading) are sufficient to make the program cost effective, however these gains are nevertheless quite small, especially when compared with the initial hopes for the program.

a more unified system of teacher support and development. My concern in this essay is not to further evaluate the program but rather to attempt to understand more about why it looks the way it does and how the design choices and constraints made at various levels of the system might help to explain some of the outcomes described above.

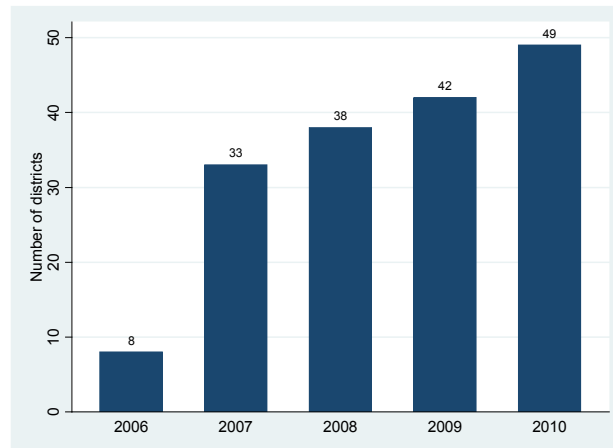


Figure 3.2: Traditional School Districts in Q Comp by Year

B. Legislation

As laid out in Minnesota statute, the Q Comp legislation was based loosely on the design of the Teacher Advancement Program. It gives MDE the power to grant state funding to traditional and charter districts for program designs that include:

- Salary schedule reform ensuring that at least 60% of any teacher pay increases are based on teacher performance. Teacher performance benchmarks can include schoolwide achievement gains, measures of student achievement at the teacher level, and/or teacher evaluation results.
- Career advancement opportunities for teachers that allow teachers to retain primary roles in student instruction.

- Site-based professional development, led by master or mentor teachers, and taking place during the school day.
- An educational improvement plan that elaborates on data collection mechanisms, links the above elements of the Q Comp program to specific school performance goals and ties these elements together into a “rigorous professional development system.”
- Joint agreement on the plan from both the local superintendent and the “exclusive bargaining representative of the teachers” (e.g. the local union leader in the case of traditional districts).

This legislation is notable both for what it does and what it does not do. On the one hand, it empowers MDE to develop a robust system of regulation for participating districts around teacher advancement, evaluation, and professional development, all with the goal of increasing the focus on teacher performance. And it calls for districts to clearly integrate these elements rather than picking and choosing among the parts. These points serve as strong evidence for the statement repeatedly echoed by those involved that the bill’s passage represented a genuine attempt to move beyond political maneuvering and into genuine school improvement. At a time when most states and most Minnesota districts were solidly entrenched within the steps and lanes system and professional development was almost entirely unregulated, the Q Comp requirements forced a real break with the past for any district that hoped to win its funding.

In addition, Q Comp was not simply a pay-for-performance bill cloaked in the language of a teacher improvement system. The Q Comp legislation looked beyond performance compensation and quite thoroughly dictated an ambitious set of supporting structures that districts were required to put in place alongside their new compensation system. Notwithstanding

the title of Alternative Teacher Professional Pay System, only one paragraph of the relevant statute actually concerns teacher pay; the rest of the legislation specifies the pieces related to professional development, evaluation, and teacher leadership that had to be included in district plans. Districts would be required to set up a structure where teams of teacher-leaders led regularly meeting professional development groups at the school level, supported each other through peer observation, and mentored new and struggling teachers. A prescribed absence of quotas on the number of performance raises was meant to ensure that teachers would collaborate in these realms and not feel the need to compete for a fixed sum of money.

To support the coherence of the structure, the legislation gave MDE the power to turn down applications that had all the individual pieces that went into Q Comp but did not tie these together into an integrated whole. To this end, the legislation required districts to submit an educational improvement plan that set definite student performance goals and benchmarks and describe the ways that the Q Comp professional development structure supports these goals. Steven Kelley (D-F-L, Hennepin), who authored the Senate bill, noted that: “One of the major strengths [of Q Comp] was the effort to align a school strategy, staff development, and evaluation of teachers so that they were all pointing in the same direction. A lot of staff development dollars had been spent on individual teacher interests that may not have been connected to the district goals.” Under the terms of the legislation, districts were expected to lay out the shape of their professional development programs within these plans and indeed were advised in the legislation that the extra state money mandated for teacher professional development at the time (2 percent of basic revenue) should not be used separately but should be combined with Q Comp money for implementation of the new system.

Yet while the legislation is quite wide-ranging, it is also deliberately ambiguous on several key points. Considerable debate within the legislature did not lead to any resolve in the law itself over the major issue of how teacher performance should be measured. Such a decision reflected the greater context where, even under the centralizing influence of No Child Left Behind and the testing system it upheld, individuals of both parties still held grave doubts about the ability of state tests to “get it right” when it came to gauging teacher success. Rather than prescribing a set of measures or even advocating the inclusion of a balanced slate of performance goals, legislators overcame their deadlock by simply giving the nod to all the possibilities. Student test results, for instance, are noted as crucial in the bill, but their uses are never fully dictated. Districts are not held to any particular test, to any way of prioritizing different classroom measurements, or even to how to balance classroom and school-level goal-setting.

This lack of prescription from the legislature extended beyond performance measurement. Indeed, many of Q Comp’s key program design decisions get delegated from the state to the district level to be decided by a team of local administrators and union representatives. Although the Minnesota Department of Education (MDE) vets applications and can withdraw funding in cases of extreme rule-breaking (this has happened once in the history of the program), localities have the responsibility of shaping the program itself, choosing which teacher performance benchmarks to measure, what teacher evaluation system to use, what teacher leadership structure can best support these processes, and they bear the entire burden of implementation.

Some of the imprecision within the Q Comp bill is of the type that can be found in any state legislative action. As Barb Sykora (R-Hennepin), the legislator who carried the Governor’s bill in the House, pointed out: “We would never have gotten a bill passed if we were terribly

specific.” But it is also the product of an educational governance setup particular to Minnesota. Unlike in many states where the state department of education is led by an elected board and/or a staff of civil servants, the core leadership of Minnesota’s department of education is appointed by the governor. Thus, a legislative decision to create a powerful and centralized program housed in MDE would have meant handing the program over to the governor, something that the Democratic senate was unwilling to do, especially after a recent and bitter battle over the governor’s initial Commissioner of Education appointment.³ Similarly, union officials in such a strongly unionized state had obvious stakes in locally controlled and locally negotiated policies. Indeed, in its initial statement of alternative compensation principles in 2000, Education Minnesota had argued that there could be “no one plan” but that plans should be negotiated through collective bargaining in order to be “flexible and structured for the context in which they will be implemented” (Education Minnesota, 2000). The result, for better or worse, as several legislators acknowledge, was that the legislature came together around a bill that would create multiple experiments in compensation reform rather than a single model throughout the state.

D. Implementation from within the Department of Education

Because the Q Comp legislation allowed considerable room for interpretation, the Minnesota Department of Education played a crucial role in determining the ultimate shape of the program. In the early years of program design, officials in the state department faced a difficult set of choices: How to implement a program that came with high expectations and

³ There seems to have been some possibility that MDE would be given fewer enforcement powers. Chas Anderson recalls a quotation from a superintendent being passed around negotiations that said that he planned on using Q Comp money for other district purposes regardless of the law’s intent since he did not think that MDE would have the power to recall the funds. As a result, language was inserted into the statute that gave MDE this power of recall.

political hopes but little clear legislative direction from within a state agency that had limited capabilities to take on instructional reform?

I find two patterns in the way the agency implemented Q Comp. First, MDE engaged in repeated political tussles with districts and the teachers union about the exact types of pay reforms that would be required by Q Comp adopters. MDE eventually created a system that was quite clear in its requirements but that allowed districts to get by with relatively small and non-disruptive changes to teachers' pay incentives.

Second, MDE coped with the difficulty of strained resources and ambiguous directives by centralizing many of the program's structural elements while leaving ideas about teaching to be worked out at the district level. This has given MDE an important role in a program that might otherwise have passed out of state hands entirely, but it did little to constrain variation across subunits or to encourage widespread learning about program success.

When Q Comp passed in special summer session, the Department of Education was immediately tasked with its implementation. As former Deputy Commissioner of MDE Chas Anderson described the situation: "The bill passed on July 17th. We were told to implement starting July 1st. We started off 17 days behind." The Department of Education, which was already struggling to meet the relatively new demands of NCLB, had little experience supporting districts in the types of teacher improvement activities defined in the new policy and few resources to devote to the problem. To make matters worse, the legislation made no provision for any administrative funding to accompany the program, leaving MDE officials scrambling both to restructure the department and its existing resources to discharge their new duties and to determine what the program should actually look like at ground level.

“Looking back,” recalled Former Commissioner Alice Seagren, “we should have had money set aside to do the administration side of [Q Comp]. We were always short of people.” Evidence for this statement can be found in the fact that Linda Trevorow, who served as initial director of the program, was not even on the official staff of MDE but rather on loan from the Minneapolis School district where local experiments with pay reform had given her the most practical experience with such programs of anyone in the building. Eventually, program control passed over to a newly created school improvement division of MDE under the leadership of the newly hired Pat King. “As we were designing and implementing Q Comp, we were designing and implementing a school improvement division,” King explained. “Given the constraints of hiring in a state agency – one week you have a hiring freeze, the next week you don’t – we were very limited on pay. School districts can pay a whole lot better than we can. It was very difficult to develop a division that had the expertise.”

King’s description captures a situation that is not atypical for many state departments of education. Indeed, the accelerated implementation timetable around Q Comp and lack of available personnel matches the current features of several states that made pledges to implement TISs as part of their Race to the Top applications. In Minnesota, this meant that the department had to quickly prioritize its goals for the program and focus in areas where it felt it had the most leverage.

As noted, Q Comp’s legislation required districts to change pay schedules to reflect teacher performance that might be measured in several different ways, and MDE played a central role in defining the rules for pay reform. At the same time, districts tended to look on the program as an opportunity to augment teacher salaries across the board, a point that I explore

further in next chapter of this dissertation. This meant that MDE faced considerable resistance to its move to establish a firm set of teacher pay guidelines.

The greatest conflict took place in 2007 when Sauk Centre school district sued MDE for denying the district's application to join the program. Sauk Centre had proposed a Q Comp program that retained teacher pay raises based on its traditional steps and lanes system but added on bonuses of \$100 per teacher if students in the district increased proficiency on the annual state tests compared with the previous year as well as a \$800 bonus for attending staff learning committee meetings. MDE rejected the proposal for not going far enough in reforming the traditional steps and lanes salary system, at which point the dispute moved first to union negotiators and then into the courts. When the Minnesota Court of Appeals ruled in favor of MDE, the decision offered vindication for the Department's approach and allowed it to further clarify the legislative statement that a percentage of salary increases must reflect teacher performance. MDE now requires that the teacher salary schedules that districts put together offer rewards for success in all three categories suggested by the legislation, including school or district-wide testing, classroom-level goal-setting, and evaluation results.

Yet this outcome, which several staff members at MDE referred to as crucial for the overall rigor of the program, still left districts with considerable freedom to minimize the salary uncertainty that teachers would face. Most districts choose to offer performance pay in the form of small bonus payments divided so that teachers receive minor payments for meeting school testing goals, minor payments for completing the evaluation process, and minor payments for meeting classroom or team-level goals. Importantly, in many districts, classroom or team-level goals have tended to be based less on direct measures of achievement than on the teachers' participation in a goal-setting and/or portfolio-making process. This means that for most teachers

in Q Comp districts, the payouts they face for student outcomes are quite low, although this does not necessarily mean that the program could not have caused important shifts in teachers' professional incentives, a point I take up in Section V of the paper.

While fighting political battles over the meaning of performance incentives, MDE also worked to bring clarity to the other provisions in the Q Comp legislation. At its core, MDE's strategy took advantage of its strengths as an agency whose primary function has been to enforce compliance with various government regulations. The Department has come to serve as Q Comp's gatekeeper, working exhaustively with applicants to ensure that program designs as specified in district applications fit within a code of requirements that have been developed by staff members. Because the legislation was least ambiguous when it came to statements listing the elements that must be included in district plans, officials in the early years of the program parsed the words of the legislation to generate a set of understandings that they eventually formalized into a series of training seminars, application guides, and "frequently asked questions." For instance, the state now specifies that the term "job-embedded professional development" requires meetings at least every two weeks and that teacher evaluations must include at least three teacher observations with at least two separate observers. Similarly, it requires that teachers' professional development meetings would need to average at least 50 minutes a week and be defined around clear, data-based goals. Thus, after receiving a district proposal, the department engages in lengthy dialogue with each applicant and in almost all cases suggests substantial revisions before the final approval process.

This attention to measurable standards carries through post-application as well. After entering the program, a district's central engagement with MDE takes place in a yearly goal-setting process whereby the district's local Q Comp leadership puts together a set of district and

school testing goals and needs to win approval from MDE before moving forward with funding for the following year. Over time, MDE has used this process to increasingly standardize the performance measurement and the rigor of the goal-setting associated with Q Comp. For instance, in the early years of the program, a number of schools defined goals using quarterly assessments put together by the not-for-profit Northwest Evaluation Association, since these formative assessments provided immediate feedback on student progress. Now, MDE requires that the goals explicitly refer to the Minnesota Comprehensive Assessment and mandates that the goals are integrated with school AYP requirements. These testing goals then become the basis for districts to award the component of teacher performance rewards that are based on school and district achievement.

What gets left out from MDE's bureaucratic oversight is the kind of long-term programmatic support that might build knowledge within districts about how to genuinely support teachers or develop their Q Comp programs over time. For example, the state recently created a rubric to assess districts Q Comp implementation. Yet due to staffing constraints that have been exacerbated as more and more districts join Q Comp, the Department has chosen to remain on the outside of this process, passing the rubric to district leaders for self- and peer-evaluations. Equally importantly, the rubric is neutral on topics like what should be covered during professional development sessions or what should be the focus of mentor relationships. Instead, when describing expectations for the job-embedded professional development component of Q Comp, the rubric is concerned with standards such as the amount of time per week that learning teams meet and the documentation that these meetings must produce. Districts and teacher-leaders are left to themselves to determine the content of the program in the form of

mentoring support, professional development curriculum, and teacher evaluation and observation materials.⁴

MDE's efforts to set standardized requirements around program structure also left the agency with few extra resources to devote to the problem of program learning and development. While members of the agency spoke about the need to promote best-practice sharing across districts and to create systems that would generate new knowledge about program design, little such work actually appears to have taken place. There is little evidence that either MDE or most of the districts involved viewed this program as an opportunity for the kind of continuous improvement process that might have increased program success over time.

In part, MDE's choice to remain detached from many of the substantive elements of Q Comp is the result of conflicts like those in Sauk Centre. While MDE won the Sauk Centre lawsuit, the furor over the case – as well as the atmosphere created by several earlier legislative hearings and complaints about MDE's enforcement of the law – left MDE wary of stretching its powers too far. But it is also a function of what is possible in a fragmented system from the state's vantage point. As state officials see it, one of Q Comp's major strengths is its focus on site-based professional development, where teacher leaders chosen at the site level train other teachers on relevant subjects. The innovation combats the traditional teacher complaint that many professional development attempts are so infrequent and removed from the classroom that they cannot be applied in a meaningful way. But it places an enormous responsibility on teacher-leaders to generate teacher improvement without offering a clear vision for how this might take place.

One way of understanding Q Comp implementation at the state level in Minnesota is to see a classic tension between centralization and decentralization. Faced with a limited set of

⁴ I explore district responses to these questions in the following chapter of this dissertation.

resources and expertise, MDE chose to standardize certain aspects of Q Comp while delegating others to district subunits. However, what is notable here is the way that MDE's choices – to centralize the structural aspects of the program while offering little guidance around teaching – demonstrate a common pattern for state educational interventions even within the bounds of a new type of state program (see, for instance, Berman & McLaughlin, 1977). Little else would have been possible within the bounds of the state system as a whole, and it is difficult to imagine that the program would have been more successful had MDE officials instead tried to use the agency's centralizing authority to promote their own views on how teaching. But if the intent of a state TIS is to create a consistent system for improving instruction across districts, it is nevertheless problematic for such programs that a state might find little place to engage in meaningful instructional guidance. In particular, this situation is likely to promote variability in an educational system where instructional quality is already highly variable. MDE's choices around Q Comp implementation shuttled the responsibility for instructional guidance passed from the state to the districts and also, less predictably, to the teachers union as I describe in the following section.

E. Union Role

Alongside MDE and the districts themselves, the teachers' union represents the final major player in the Q Comp story. Not only is the union, known as Education Minnesota (EdMinn), particularly powerful since it represents the combined force of both the AFT and the NEA, but it is highly representative of the state teacher corps. Indeed, around 99% of teachers in Minnesota are currently union members. As a third party with a complex relationship to the demands of teaching improvement systems, the union brought an unpredictable dynamic to the Q

Comp reform movement acting as both partner and antagonist at different points in time. But through both its state and local branches, it also temporarily served as a crucial and somewhat surprising intermediary in the process, offering supplemental services and knowledge that were not or could not be provided by MDE.

As noted previously, the state union spent several years in anticipation of Q Comp educating its members and defining its own stance around alternative compensation. Lobbyists from EdMinn played a substantial role in defining the possibilities for the legislation and in shaping the Senate version of the bill. Organization officials say they still had some reservations when the bill passed in 2005, especially because Governor Pawlenty seemed to view the law as taking a harsh stance toward struggling teachers. “Obviously the legislation was a political agenda item for Governor Pawlenty,” EdMinn Professional Pay Systems Coordinator Randi Kirchner observed. “He saw it as more punitive.” However, many of the principles that EdMinn had set out in 2000 were captured by the new legislation, including district-level flexibility, strong support for professional development, and the pledge that the new system would not reduce base pay for any teacher. Most importantly, the Q Comp plans put together in each district required the agreement of the local union representative, giving the union an important part to play in program design. Together, these factors helped win tentative support from EdMinn in the legislation’s initial months.

The traditional narrative around teacher pay reform pits policymakers against unions, and there is speculation among a number of Q Comp opponents that Governor Pawlenty meant the policy as a first step in a larger attack on the power of teachers unions. However, the compromise legislation that Q Comp represented in many ways had the opposite effect. Minnesota’s legislation actually significantly expanded the union role in many ways by moving

state and local branches to combine traditional salary negotiations with larger policy discussions. Under the new rules, all Q Comp policies including professional development structures, teacher evaluation designs, and mentor teacher roles must be hashed out between district administration and the local union in collective bargaining negotiations.⁵

“Education policy became a mandatory subject of bargaining in Minnesota,” said Randi Kirchner, the EdMinn Professional Pay Systems Coordinator at the time Q Comp was passed. “Under traditional contract negotiations, testing, professional development, etc. are managerial. But now, for the first time, teachers had an equal say at the table about these things.” Former President Judy Shaubach expressed a similar idea: “Educational policy was something that prior to this was never part of what the union was allowed to bargain... This was a way for the union to have a very significant role in thinking about how to improve teaching.”

To support this role, EdMinn devoted a significant chunk of its resources toward the new program. In particular, the organization began to hold training sessions and work directly with districts to help them interpret the legislation and submit their plans to MDE for approval. Unsurprisingly, this process provoked a number of clashes between MDE and union representatives over areas where interpretations differed and at times left some bitter feeling between both parties. But it also provided an opportunity for the union to become an important source of expanding district knowledge.

The union also worked in collaboration with local branches to help build resources for teacher professional development within the bounds of the Q Comp program. For instance, several districts when faced with the question of how to expand internal capacity for professional

⁵ Interestingly, this collective bargaining requirement also appears to hold some of the blame for Q Comp’s slow growth across the state. Not only did some districts not apply to Q Comp because they not find the common ground between the administration and the union to create a joint application, but other districts turned away from the program for fear that the state funds would dry up and the district would be forced to cover the costs since the program had been enshrined in its local collective bargaining agreement.

development turned to the Education Research and Dissemination program run by the AFT that trains teachers to instruct their colleagues in research-based teaching techniques. These relationships opened up new opportunities for collaboration between union and district administration and proved a fruitful source of help when districts needed to find resources that were not being provided by MDE.

However, even as the union seemed poised to step into a new role of supporting partner alongside MDE, the collaboration began to fracture over differing interpretations of the ambiguous Q Comp legislation. To union leadership, the requirement to “reform the steps and lanes salary schedule” suggested that districts should be allowed to make only minor additions to the schedule, including those that distributed Q Comp money evenly among all teachers who took part in the new professional development. MDE took the stance that pay reforms under Q Comp needed to be partially based on performance, although its record suffered from the fact that the agency had been fairly inconsistent about exactly what it wanted during the first year of implementation. As noted earlier, the dispute culminated in the Sauk Centre lawsuit and a sense among union officials that they had been shouldered aside by a power-hungry state agency. A leadership change in the union further fractured the growing relationships between officials in the state and the union who had grown comfortable working together around Q Comp.

Over time, the state union largely disengaged from taking an active role in Q Comp policy management, even as local union branches continued to broker their district-level Q Comp contracts. The result was that district programs lost out on what might have been a crucial source of knowledge and support on how to develop and support teacher learning. As Shaubach explains: “When we started to have a lot of these challenges over how [Q Comp] was interpreted, it really put a damper in sustaining this long term... If there had been a better

relationship or a better partnership between the department and the union where we could have together provided more support and oversight, that certainly would have helped districts to sustain what they started or to help new districts engage in this process.”

V. Interpreting What Happened in Minnesota

When the Minnesota legislature began the process of reforming teacher pay, career pathways, and professional development in 2005, it embarked upon a highly ambitious enterprise with little precedent at the state level. In effect, legislators expected a cash- and resource-weak department of education to transform teacher human capital policies within a state where nearly every district paid teachers based on a straightforward steps and lanes schedule and drew few explicit connections between pay, evaluation, professional development, and performance. Moreover, this transformation was to take place in collaboration with a powerful teachers union that had declared opposition to programs that significantly increased teachers’ salary uncertainty or decreased policy flexibility at the district level. That lawmakers in collaboration with the department of education were able to put together a program that still endures today rather than suffering an early demise in the midst of repeated budget crises and power shifts speaks to the political success of the endeavor.

But what did the maneuvering around design and implementation mean for the programming itself? At the beginning of this essay, I suggested that TISs might drive teacher improvement by increasing the alignment between teachers’ incentives and their opportunities to learn how to reach these incentives. Yet I also noted the difficulty of adopting such programs when the current structure of the system reduces the likelihood that any centralized state agency

can provide focused and coherent instructional guidance. The Minnesota case both bears out these observations and offers further insights.

On the incentive and measurement side of the TIS, there has been an interesting progression. At the legislative level, lawmakers in 2005 proved unable or unwilling to commit to a specified system of teacher performance measurement. The law that they passed required districts to revise teacher salary schedules in order to include performance measurement, but it offered little guidance as to how this might be done effectively.

The buck thus passed to MDE, which might have simply pushed it on through to local district authorities. Indeed, in the initial years of the program, the department gave districts considerable autonomy to define every aspect of their teacher performance system. But over time, MDE used its role as the program gatekeeper to define and standardize several of the program components relating to teacher incentives. In this, MDE used the tools at its immediate disposal, including state tests and compliance frameworks. This meant creating salary schedule requirements defining the extent to which the new schedules needed to include measures of teacher performance and pushing districts to make significant use of statewide tests and school-level AYP goals in these measurements. The program that developed out of this process was in some ways remarkably specified for such a state-level endeavor. But these conditions did not necessarily produce a program that would produce widespread instructional improvement.

One design issue concerns the strength of teachers' performance incentives. As noted, notwithstanding the rhetoric around Q Comp as a performance pay program, most teachers in most districts faced little pay uncertainty connected with student outcomes.

Such a conclusion is not necessarily a program weakness; indeed, there are reasons to believe that greater pay uncertainty could have promoted unfortunate outcomes like an increase

in attempts to “game the system” (Rothstein, 2008). Moreover, I don’t mean to imply that the pay reforms that districts put in place as a result of Q Comp are meaningless. Even if the amount of money based on student performance is small, the extra bonus might serve as an important signal for what the district values, and there have been no studies of pay incentives in education that show that larger bonuses have correspondingly larger effects.

In their seminal piece on the use of merit pay in education, Murnane and Cohen (1986) suggested that most merit pay programs that survive over time tend to look quite similar to those that Q Comp has created, and that these programs can serve important purposes within the district. The authors argue that most such programs can be found to employ a variety of techniques: “Extra pay for extra work; make everyone feel special; make the program inconspicuous, and legitimation through participation” (p. 12). The point of these programs, the authors find, is less to reward the most effective teachers than it is to support a system that allows greater teacher income differentiation and that encourages dialogue between teachers and administrators around teacher evaluation. In the context of Q Comp, such a system might have the desired effect of promoting greater awareness of and communication about instructional technique across classrooms and it could encourage teachers to renew their focus on student achievement.

At the same time, the weakness of teachers’ pay incentives raises questions about the extent to which the program shifted teachers’ focus in the direction of student achievement as policymakers intended. At the very least, the relative weakness of the pay reforms in Q Comp shifted much of the burden for program effects onto the other elements of Q Comp. Moreover, the pay reforms that developed through Q Comp in Minnesota create an important contrast between this and some of the newer Race-to-the-Top programs such as the system in Tennessee,

where pay based on the classroom-level student performance represents a significantly larger element of the program.

With respect to the non-pay-related elements of Q Comp, MDE's choices had both positive and negative consequences. The state's efforts at program implementation established a workable program that required significant commitment from adopting districts. Moreover, by establishing initial statewide standards for the system, MDE laid the framework for a blanket teacher evaluation requirement that the state plans to put in place by 2016. However, while MDE's process created the skeleton of a teacher improvement system, it did not add flesh to the bones. Few of MDE's decisions were designed to directly enhance teachers' knowledge of how to meet their specific program goals or, in the jargon of incentive studies, their cognitive capital.

Districts that won approval from MDE to join Q Comp were required to create regularly occurring venues for teachers to share ideas, and they were required to create a system in which each teacher was observed at least three times per year and received rapid feedback according to a standardized rubric. Moreover, they were required to document that these things were taking place and provide MDE with this documentation. These records serve as an important record of the policy's impact. There is considerable evidence that districts transformed professional development practices in response to Q Comp in ways that allowed more knowledge sharing to take place between teachers.

Such changes are not negligible. Indeed, there is some reason to believe that these reforms alone might positively influence both teacher professionalism and student learning. For instance, a recent study of teacher evaluation in Cincinnati found that the process of undergoing repeated teacher evaluation seemed to improve outcomes for a set of several hundred mid-career teachers (Taylor & Tyler, 2011). Moreover, an extensive literature on teacher professional

learning communities (PLCs) argues that PLCs can substantially influence teacher culture by creating a forum for sharing practice-based ideas (Vescio, Ross, & Adams, 2008; Coburn & Stein, 2006).

But in the Q Comp model, the responsibility for promoting meaningful instructional change largely falls upon local leadership. Program districts had to make their own decisions about what teacher observers should look for, about what PLCs should discuss, and about how to interpret various sources of teacher and student data. Because Q Comp leadership varied considerably in its capacity to take charge of these processes, the result was substantial unevenness in perhaps the central piece determining programmatic success or failure.

This did not mean that districts never turned to other sources for some of the components associated with their Q Comp plans. For instance, since MDE demanded formal observation rubrics to be submitted with Q Comp applications, many districts used Charlotte Danielson's materials for teacher evaluation, and, indeed, MDE often pushed for this model in initial discussions. But most districts adopted a piecemeal approach to bringing in these external aids, often, for instance, simply appropriating a piece of the Danielson rubric for teacher observations but not sending teachers out for a full course of training and implementation support. To some degree, this strategy can be traced back to the budget squeeze that caused many districts to turn to the state program in the first place; since districts often hoped to use Q Comp funds as a substitute for other increases in teacher pay, they were loathe to spend too much of the money on outside training. Moreover, teachers had little extra time to devote to overly substantial training since they were already using their extra time for the new forms of professional development taking place within the district.

This structure has several disadvantages. Q Comp cuts down on teachers' isolation but does not necessarily add to their knowledge base. If some teachers in the district have clear ideas about what works, they can transmit them to others through the new professional development structure, but if there is less certainty or if certainty is misplaced, Q Comp provides little. Meanwhile, the new requirements also add an entirely new layer of bureaucracy to a system that was not short on bureaucracy to begin with. Indeed, one way of interpreting Q Comp is as another contributor to the fragmentation of governance and instructional guidance that I described at the outset of this paper. Rather than clarifying the system, Q Comp simply superimposes an intermediary level of instructional leadership on top of what is already taking place, tasking teachers within each newly adopting district with developing their own systems of professional guidance and feedback. This in turn further contributes to the already present sense that each individual subunit within a state—schools, districts, teaching teams, etc.—must reinvent the wheel around instructional guidance. Moreover, it can promote further inequality in an already unequal system if some units due to initial advantage are able to more successfully draw in resources to make this task more manageable.

This analysis should not be read to imply that Q Comp would have been a more successful program if decisions about instruction had been further centralized at the state level. Indeed, there is little evidence to suggest that the state possessed the resources, the expertise, or the policy tools that would have allowed it to play a greater role in instructional guidance. But at the same time the lack of guidance or knowledge-sharing at each level of the system contributed to a program with uneven results and little improvement over time.

What might have helped to mitigate these factors and contributed to greater Q Comp success? One possibility would have been an entirely different approach from within MDE. To

some degree, MDE's hands-off approach to district professional development makes sense for a program that always had a strong local focus and was intended to provide districts with money to set up their own context-specific teacher development plans. Yet retaining local autonomy does not necessarily require complete non-interference. It would be possible to imagine a situation where MDE had decided that supporting Q Comp meant providing substantial support to districts around particular types of teaching and/or learning. Such support might still leave districts a great deal of control over their individual goals while offering a stronger array of resources around meeting these goals. Yet offering such support would also have required a great deal more from MDE in terms of expertise, staff, and funding than the organization is equipped to provide. As more districts entered the program, each with its own individual needs regarding curriculum and teacher training, this task would have continued to grow, making MDE's job ever broader and the prospect ever more daunting.

Another answer might be to look to outside organizations to supplement MDE's influence. One of the more interesting developments in Minnesota was the way the teachers' union sometimes stepped into the breach to provide supplementary training and support alongside MDE. Indeed, the setup of the original legislation, which required the collective sign-off of superintendent and local union representative helped to draw the state organization into the partnership. Initially, when the union devoted an entire division and much of its own funding to helping districts with program application and setup, it appeared that Minnesota might be poised to forge an unprecedented collaboration around Q Comp with the union relinquishing its typical focus on teacher salaries and contracts for a far broader role supporting teaching and learning. This still appears to be the case within a handful of districts where the district-union partnership forged by Q Comp has leaned heavily on union know-how and resources at the national level to

bolster its professional development program. However, at the state level, there has been far less progress. Repeated conflicts between union and MDE officials over interpretation of the Q Comp legislation created a lasting antagonism, and, after the Sauk Centre lawsuit, the union slowly began to withdraw its resources from Q Comp. At this point, there is no one in state union whose job is solely devoted to Q Comp support, and indeed the union recently took a formal stance against the program in the state's application for Race to the Top funding.

Another place that the state might have looked for assistance was to the many private and non-profit organizations that provide district-level teacher support. To some extent, such a process developed organically, with some districts contracting for training with Charlotte Danielson's organization for help on evaluation rubrics and others turning to similar groups for different aspects of their Q Comp plans, yet as noted, the adoption was often piecemeal and only related to individual aspects of the improvement system. Greater centralized support for district partnerships with these organizations might have offered districts a source of more robust guidance on program development. This would not have reduced the difficulty of providing unified instructional support within a fragmented system, but it could have offloaded the responsibility to organizations that face fewer constraints than the state department. Several researchers have described models of public-private cooperation, where semi-autonomous organizations are able to create models of improvement that would have been difficult to develop from deeper within the system (Rosenberg, 2012; Glazer & Peurach, in press; Foley, 2010).

Another modest change that might have had positive effects on districts' abilities to make positive use of the Q Comp structure would have been for the state to do more to promote cooperation across districts. Most state officials envisioned a process within Q Comp where districts would develop their own plans and improve them over time based on what worked best

in other Q Comp districts across the state. To encourage such idea sharing, the state posted summaries of district plans online and adopted a program review process where Q Comp leaders in one Q Comp district would visit other Q Comp districts and comment on their operations. However, there was little support for a more organized process of cross-district learning and development. The point here is not that Q Comp did not include a number of positive elements, but that it could have been more successful if it included more institutional systems to support the development of greater knowledge about program design over time. Without the state acting as a conduit for this process, the contacts that occurred tended to be haphazard and the sharing of best practices anecdotal rather than based on strong evidence. Indeed, the only significant place where district idea-sharing seemed to take place was during the application stage when newly applying districts would look to previously accepted applications posted online on MDE's website for hints on the language that would prove acceptable to the state agency.

VI. Conclusion

In the next several years, a number of states will redraw their teacher development policies with the goal of strengthening teacher evaluation, allowing greater pay differentiation, and enhancing professional learning opportunities. A central principle supporting these changes is that they will induce substantial improvement among current teachers by improving the alignment between teachers' goals and their knowledge of how to meet these goals. Thus, the reforms are meant to be seen as coherent improvement systems rather than disconnected policies.

I have argued that Minnesota offers an important window into these upcoming developments. There are several differences between more recent plans and the reforms that Minnesota began in 2005, most importantly that the newer plans are often meant to be eventually

implemented in all districts across the state rather than only in districts that apply for state funding. However, there is also considerable overlap, as state departments of education take on the responsibility for creating entirely new standards and regulations around district human capital practice.

My study of the Minnesota case study identified several challenges around designing and implementing TISs from the state level. On the one hand, states are attempting to create increasingly coherent supports for teacher learning, yet they possess few of the tools around instructional guidance to make this possible. To make matters more difficult, the central agencies charged with this task are often underfunded and working on accelerated timetables that leave little room for prolonged program design or engagement with districts. That MDE was able to put together a long-lasting program and sensibly designed program under such circumstances is a tribute to the department. However, the department's decision to focus on program process over content made its contribution to instructional practice and therefore student achievement uncertain and likely contributed to the variance in program effects across districts.

As other states take on these challenges, one might hope that the legislative branch could solve some of these problems. In the case of Minnesota, changes to the legislation might have helped. For instance, more clarity about what was meant by the measurement of teacher performance could have reduced the scope of MDE's initial battles with districts and the teachers union. Similarly, if MDE had been provided with a substantial budget for administering Q Comp, it might have been able to create a team of staff solely devoted to working with districts around their Q Comp teacher development policies. However, these changes would not have solved the central issue of how to create mechanisms for teacher learning within a system that offers few centralized supports for such a process. This issue is not likely to be worked out by

the legislative branch but rather in the lower-level decisions around implementation. The ways these decisions get made, and the role that state departments come to play, are likely to prove crucial to the long-term success of these reforms.

With momentum around such policy changes continuing to grow, my account suggests a number of issues that merit further investigation. First, it will be important to determine how committed state departments actually are to encouraging the teacher learning that makes up one half of the TIS structure. While much of the current rhetoric around the new teacher evaluation and promotion policies contains reference to the need for supporting teacher professional development, it is possible that this rhetoric might be dropped in states other than Minnesota once the policies make their way into law. Second, if teacher learning is to remain a crucial piece of these policies, the policies must be accompanied by further attempts to define the domains in which this learning might occur. While it makes sense to try to develop teachers' cognitive capital to meet new incentives, states will continue to have little power to influence this development unless they create a better set of tools for developing instruction across schools and districts. Third, as these systems develop further in other states, it will be useful for researchers to find new ways to consider the interaction between teacher learning and incentives. At the state and district levels, researchers might ask a variety of questions about the alignment between teachers' professional development and their professional goals that could help to shed light on the potential of these systems to influence classroom practice and student achievement.

REFERENCES

- Berman, P. and McLaughlin, M. (1977). *Federal programs supporting educational change*. Santa Monica, CA: RAND.
- Bonner, S. E., & Sprinkle, G. B. (2002). "The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research." *Accounting, Organizations and Society*. 27: 303-345.
- Camerer, C.F., and Hogarth, R.M. (1999). "The effects of financial incentives in experiments: A review and capital-labor-production framework." *Journal of Risk and Uncertainty*. 19(1-3): 7-42.
- Coburn, C. E. and Stein, M. K. (2006). "Communities of practice theory and the role of teacher professional community in policy implementation." In M. E. Honig (Ed) *New directions in education policy implementation: Confronting complexity*. Albany: State University of New York Press.
- Cohen, D. K., (2011) *Teaching and its predicaments*. Cambridge, MA: Harvard University Press.
- Cohen, D. K. and Hill, H. C. (2001). *Learning policy*. New Haven, CT: Yale University Press.
- Cohen, D. K. and Moffitt, S. L. (2009). *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Cambridge, MA: Harvard University Press.
- Cohen, David, K. and Spillane, James, P. (1993). "Policy and practice: The relations between governance and instruction." *Review of Research in Education*. 18: 3-49.
- Corbin, J. and Strauss, A. (2008). *Basics of qualitative research 3e*. Thousand Oaks, CA: Sage Publications.
- Education Minnesota. (2000). "Professional Advocacy Committee Recommendations – Alternative Compensation." Manuscript provided by Education Minnesota.
- Elmore, Richard, F. (1996). "Getting to scale with good educational practice." *Harvard Educational Review*. 66(1): 1-27.
- Elmore, Richard F., (2004) *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard Education Press.
- Foley, Eileen, M. (2010). "Approaches of Gates Foundation-funded intermediary organizations to structuring and supporting small high schools in New York City." Policy Study Associates. Accessed online at <http://www.policystudies.com/studies/?id=1> on April 4, 2012.

- Glazer, J.L. and Peurach, D.J. (In press). "School improvement networks as a strategy for large-scale reform: The role of educational environments." *Education Policy*.
- Glazerman, S., and Seifullah, A. (2010). "An evaluation of the Teacher Advancement Program (TAP) in Chicago: Year two impact report." Washington: Mathematica Policy Research. May 17.
- Hezel Associates. (2009). *Quality compensation for teachers summative evaluation*. Accessed online at <http://archive.leg.state.mn.us/docs/2009/other/090321.pdf> on April 4, 2012.
- Lankford, H, Loeb, S., and Wyckoff, J. (2002). "Teacher sorting and the plight of urban schools: A descriptive analysis." *Educational Evaluation and Policy Analysis*. 24(1): 37-62.
- Lortie, Dan C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- McLaughlin, M. (1987). "Learning from experience: Lessons from policy implementation." *Educational evaluation and policy analysis*. 9: 171-178.
- Merriam, Sharan, M. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: John Wiley and Sons.
- Miles, K. H., Odden, A., Feranich, M., Archibald, S., and Gallagher, A. (2004). "Inside the black box of school district spending on professional development: Lessons from comparing five urban districts." *Journal of Education Finance*. 30(1): 1-26.
- Miles, M. B. and Huberman, M. A. (1994). *Qualitative data analysis: An expanded sourcebook*. 2nd edition. Newbury Park, CA: Sage Publications.
- Murnane, R. J. and Cohen, D. K. (1986). "Merit pay and the evaluation problem: Understanding why most merit pay plans fail and a few survive." *Harvard Educational Review*. 56(1): 1-18.
- Odden, Allan. (2008). "New teacher pay structures: The compensation side of the strategic management of human capital." Consortium for Policy Research in Education. Strategic Management of Human Capital Manuscript.
- Podgursky, M. and Springer, M. (2007). "Teacher performance pay: A review" *Journal of Policy Analysis and Management*. 26(4): 909-949.
- Rothstein, R. (2008). "Holding accountability to account: How scholarship and experience in other fields inform exploration of performance incentives in education." Working Paper 2008-04. National Center on Performance Incentives.
- Sawchuk, S. (2009). TAP: More than performance pay. *Education Week*, April 1. 28(27), 25–27.

- Shulman, Lee, S. (1987). "Knowledge and teaching: Foundations of the new reform." *Harvard Educational Review*. 57(1): 161-182.
- Siemson, E., Roth, A.V., Sridhar, B. (2008). "How motivation, opportunity, and ability drive knowledge-sharing: The constraining factor model." *Journal of Operations Management*. 26: 426-445.
- Sojourner, A., West, K., Mykerezi, E. (2012). "Teacher pay reform can work: Panel data evidence from adoptions of Q Comp in Minnesota." Working Paper, University of Minnesota.
- Spillane, James, P. (2000). "Cognition and policy implementation: District policymakers and the reform of mathematics education." *Cognition and Instruction*. 18(2): 141-179.
- Spillane, James, P. and Jennings, Nancy, E. (1997). "Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective." *Teachers College Record*. 98(3): 439-481.
- Schwartz, N. L. (2012). "What can we expect from teacher improvement systems? A quantitative evaluation of Q Comp in Minnesota." University of Michigan Dissertation Essay.
- Taylor, E. S. and Tyler, J. H. (2011). "The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers." National Bureau of Education Research Working Paper Series. Working Paper 16877.
- Vescio, V., Ross, D., and Adams, A. (2008). "A review of research on the impact of professional learning communities on teaching practice and student learning." *Teaching and Teacher Education*. 24: 80-91.

CHAPTER FOUR

DEFINING INSTRUCTIONAL EXPERTISE: PATTERNS OF DISTRICT VARIATION IN Q COMP IMPLEMENTATION

I. INTRODUCTION

Recent scholarship on large-scale instructional reform places teacher and administrator learning at the center of the reform process. According to these views, successful policy implementation depends to a large extent on whether school personnel have the opportunity and the incentives to engage with new ideas in ways that change their understanding of their work and thus their on-the-job practices (Cohen & Barnes, 1993a and 1993b; Spillane & Thompson, 1997; Knapp, 1997; Cobb & Jackson, 2011). This literature adds to and complicates previous work on reform implementation that focused on the alignment between policy demands and local capacity (Pressman & Wildavsky, 1973, Berman & McLaughlin, 1977, McLaughlin, 1990). In the newer framework, capacity depends in part on local needs and abilities but also on whether local systems are set up in ways that promote genuine learning about instruction. As Spillane and Thompson (1997, p. 187) observe, if instructional reform is about learning, then district capacity can be thought of primarily as “a capacity to learn the substantive ideas at the heart of the reform and to help teachers and others within the district learn these ideas.”

A central idea here is that policy reforms depend on ideas and the ways they get defined and transmitted as they transition through the educational system. However, much of the research into district capacity for instructional reform tends to focus on the transmission side of this equation. There has been less work on how instructional expertise is defined within the

bounds of a reform program and what this means for the process of policy implementation. This research focus can be explained in part because many of the reform implementation studies have focused on similarly-structured state-based reforms in the 1980s where instructional expertise was typically supposed to be lodged in new curriculum frameworks. In contrast, recent large-scale policies aimed at classroom change have focused less on curriculum and more on improving teachers' human capital. This is particularly true of the new reforms promoted by federal programs like Race to the Top that ask states to adopt new means of teacher evaluation, tenure and/or pay, and professional development. In these "teacher improvement systems" (Schwartz, 2012), it is far less clear what instructional expertise looks like or where it is to be found. The answers to these questions are not evident in top-level policy descriptions but develop instead at the local level as district and school policy-makers hammer out the details of the program.

In this article, I investigate the ways that three similarly placed districts designed and implemented local initiatives as part of Minnesota's Q Comp program. In particular, I explore the markedly different answers that the districts reached about the sources of instructional expertise in the program, the ways these ideas were instantiated in their local designs, and the consequences that these ideas appear to have had on the operation of their programs.

Although Q Comp was created in 2005, it represents an important predecessor of the type of design that a number of states are currently scaling up as a result of the federal Race to the Top grant competition. It provides state funding for districts to implement locally developed plans that meet state requirements around teacher evaluation, performance pay, career advancement, and professional development. It has been joined by about 45 traditional and 30 charter districts, encompassing around one-third of the state's student body.

By investigating, how three early adopter districts designed and implemented local plans within state guidelines, I am able to explore an important source of program variability and understand more about the interaction between state and local policy. These case studies provide an important window into the contingent nature of the district program design process within the state's political organization. I argue that the structure of Q Comp's design at the state level left districts to play the central part in determining what teachers were meant to learn from the program and how they were meant to learn it. In each district, I find a different answer to this question. One district emphasizes peer-to-peer feedback, another focuses on the analysis of student data, and a third makes use of professional development content from an external agency. These differences seem to derive in large part from the distribution of design resources and the ways that instructional expertise had been defined within the district through previous programs and interventions. Notwithstanding other structural similarities in program design, these variations resulted in important differences in program operation. They also offered various challenges to the districts, and I explore the implications of these challenges to the design of such interventions.

II. CONCEPTUAL FRAME

Encouraged by recent research demonstrating both the importance of good teaching to students' long-term outcomes and the extent of variation in teacher effectiveness across classrooms (Chetty, 2010; Kane, Rockoff, & Staiger, 2008), many researchers and policymakers have come to view teacher instructional practice as a central point of leverage for policy reform. Instructional practice is often described as the area where "the rubber meets the road," or the

“technical core” that must change for there to be genuine school improvement (Diamond, 2007; Elmore, 1996; Meyer et al., 1992).

Recent reforms spurred by the federal Race to the Top, where states aim to change the ways that teachers are paid, promoted, evaluated, and developed, explicitly pursue instructional change through policies that tie human resource strategy to professional development programming (NCTQ, 2011). Over the next several years, the majority of states will transform their teacher workforce policies through a series of interlinked policy changes, and several including Tennessee, Delaware, and the District of Columbia have already done so. These reforms represent a substantial departure from the typical district practice that has tended to separate the various components of the teacher workforce policy into separate and non-interacting divisions (Odden, 2008; Miles et al., 2004).

In a previous piece about such programs, which I refer to as Teacher Improvement Systems, I described a theory of action where these programs might stimulate on-the-job improvement by shifting teachers’ incentives while simultaneously increasing the teachers’ ability to respond to these incentives (Schwartz, 2012). According to this theory, one reason that several other plausible instructional interventions have generally not produced meaningful achievement effects is that the interventions did not take into account the interlinked relationship between teachers’ professional rewards and their knowledge and skills.¹

The idea of complementary human capital strategies draws on ideas from fields like business economics where researchers have noted significant variation in subjects’ responses to

¹ The theory of action I describe here is not the only way of understanding teacher improvement systems. Another view is that these systems are mechanisms for changing the composition of the teaching force by pushing out ineffective teachers and rewarding effectiveness. However, while the teacher sorting piece is certainly one element of policymakers’ strategy, many of those involved in the design and implementation of these programs suggest that these they are primarily focused on raising the quality of the current teacher workforce (Jerald, 2012). According to my interview subjects, this is certainly the case for the Minnesota program that is the subject of this study.

various types of professional incentives that depends on the mix of knowledge and skills that the subjects bring to the tasks as well as their opportunities to augment this expertise (Siemson, Roth, & Sridhar, 2008; Bonner & Sprinkle, 2002; Camerer & Hogarth, 1999). Camerer and Hogarth coin the term “cognitive capital” to refer to specialized knowledge, “the variety of tricks or approaches” that directly apply to the demands of the work (p. 9). The idea of cognitive capital provides a useful frame for organizing studies of teacher improvement systems since it suggests the importance of looking beyond only the evaluative aspects of these programs to incorporate the broader goals of structured instructional improvement.

Within this frame, one way to assess the design of a teacher improvement system is to assess the program’s likelihood of building cognitive capital among the users of the system to meet the program’s stated goals. This question in turn raises a series of sub questions around the teacher learning elements of the program. To what extent does the program allow users to supplement their current knowledge and skills with new and applicable ideas? To what extent does it develop a shared knowledge base that users can call on for their particular needs? To what degree is the set of knowledge and skills that the program offers aligned with the incentives that the program provides?

To date, there has been little research on what teacher learning might look like within the bounds of such a program or how the design features of the program might be aimed to more or less strongly support knowledge-building. Understanding these features of such systems is a key piece in assessing their potential for genuine improvement.

For the set of reforms currently in development, this issue requires careful attention to implementation at the district level. Although most of the recent reform systems themselves originate with the state, many of the features most relevant to teacher development are likely to

be worked out by individual districts and schools. The state departments that have charge of the larger programs often have little inclination to enter into the direct work of instructional supervision, and they tend to be limited in their training efforts by lack of funding, personnel, and know-how (Schwartz, 2012). To a substantial extent, this is the product of the highly localized system of educational control in the United States. State agencies possess few of the tools in the form of common curricula, teaching frameworks, or power over teacher education that might enable them to offer strong guidance around instructional improvement (Cohen & Moffit, 2009). Thus, states have traditionally remained on the sidelines with respect to teacher development, and are ill-equipped in terms of expertise, funding, or capacity to take a more active role. As a consequence, districts and individual schools stand to play a major role in the enactment of several important elements of the new policies.

Previous research into several waves of large-scale instructional reform over the last half-century provides an important perspective on the difficulties of local implementation within a fragmented system of educational governance where responsibilities for funding and decision-making are divided across several layers of actors (Cohen & Spillane, 1993). This literature offers two major ideas that are highly relevant to the development of teacher improvement systems.

Much of the early work in this area focused on the problem of creating rigorous and sustainable programs in a system where the goals of top-level policy-makers were not necessarily shared by local implementers. Even in cases where the different parties shared basic goals, researchers noted, local actors faced potentially conflicting demands from their current work that could not necessarily be anticipated by initial policy statements. This research suggested that policy implementation involves a developing process of mutual adaptation driven

largely by the needs of the local side and shaped by a largely unstructured system of governance. Local opportunism and transformation are to be expected in all aspects of programming. Thus, reform initiatives often tend toward extreme variation in implementation with the majority of localities making superficial changes necessary to remain involved while avoiding significant transformation in key areas (e.g. Berman & McLaughlin, 1977; McLaughlin, 1990; Weatherly & Lipsky, 1977; Firestone & Corbett, 1988).

Later scholarship shifted the focus from concerns about local compliance to concerns about local understanding and interpretation. Several studies looking at 1980s-era curriculum reforms noted that teachers and administrators often shared similar beliefs that they had fully adopted the components of a new policy even as they differed considerably in where they ended up and even in what they thought the new policy asked for. This work suggested that implementation depended in large part on the process of engagement with new ideas and innovations, leading Spillane (1997) to describe the process of moving from policy to implementation as “a process in which teachers and others must change their minds in order to change their practice” (p. 186). This research highlighted the intellectual work of implementation and the ways that different policies allowed more or less room for this work. Unless a policy included the means for implementers to reach shared understandings about what the policy actually meant and, more specifically, about what it meant for teaching, it was unlikely to produce substantial improvement. Continued work in this vein explored the transmission of ideas through different types of policy directives and the ways that beliefs, dispositions, and previous experiences of various implementers affected the ways they made sense of these policy demands (e.g. Spillane, Reiser, & Reimer, 2002; Coburn, 2001; Jennings, 1996; Cohen & Barnes, 1993a and 1993b).

Each of these lines of thought offers useful insight into the design and development of teacher improvement systems. First, these studies suggest the importance of viewing such programs at the level of individual districts rather than as a homogenous operation, assuming that variation will be the rule rather than the exception. From this perspective, it becomes important to look for patterns in local demands on the program and the ways these patterns are instantiated in local programming. Second, the research highlights the central importance of the teacher learning process in mediating policy implementation, a point that is highly relevant given the organization of the new programs and the notion that better alignment between incentives and professional development opportunities might help build the cognitive capital to spur genuine improvement. In theory, the structure of professional development within these new systems, including things like professional learning communities and peer coaching could provide teachers with the necessary opportunities take a deep look at their own instructional practice.

At the same time, the previous research tends to assume that district policymaking about instruction takes place within the context of some set of relatively standardized policy guidelines (such as standards or new curricula) and that variation enters in because district personnel interpret these guidelines in individualistic ways. The teacher improvement systems currently in development are quite different than the round of curriculum-focused reforms that tended to be the focus of previous implementation studies. Where these earlier initiatives had the intent of imposing particular visions of teaching and learning such as new techniques for teaching critical thinking in mathematics, the current crop of reforms propose a model where the source of expertise is even less defined.

These new policies present something of a paradox. They stress the importance of teacher learning without defining what it is important to learn. Commenting on teacher preparation

methods, David K. Cohen (2010) cogently observed that “to teach is always to teach something.” Similarly, to learn is to learn something. Yet what teachers are meant to learn by dint of these new structures remains unclear or at least unspecified. Rather than favoring any particular content, the structures are meant to create room for implementers at the district and school levels to reach their own answers about what is important. Previous implementation studies have little to say about local development of expertise within the context of a larger program of reform.

This gap leaves many important questions unanswered. What strategies are districts using to encourage teacher learning? What are the relevant design features of the systems at the state and district level and how do they support the development of teacher expertise? To what extent do district strategies seem likely to support the kind of teacher development that is called for by theories about teacher improvement systems and to what extent do these strategies seem likely to generate large-scale improvement in instructional quality?

In the following sections, I explore the design and implementation of a state-directed teacher improvement system in three districts, focusing on structures for instructional learning and the effects on program operations. The basis for my study is a program called Q Comp that was adopted in Minnesota in 2005. While the program predated the Race to the Top reforms, it bears a strong resemblance. Directed at the state level, Q Comp asks districts to submit proposals according to state guidelines for proposals that make provision for teacher observation and evaluation, a teacher career ladder, and a salary schedule based partly on teacher performance.² Within each of my cases, I look to understand patterns of adaptation within the state’s guidelines and the extent to which local design choices appear consistent with the ideas about cognitive capital and teacher improvement systems sketched earlier in this section.

² The major difference between Q Comp and the more recent RTTT reforms appears to be that Q Comp was never expected to include all districts in the state while the new programs require full adoption within several years.

III. RESEARCH METHODS

A. Site Selection

This study uses a qualitative, cross-case design in order to explore district influence on Q Comp design and implementation (Patton, 2002). Three school districts serve as the central units of analysis. Because I wanted to be able to separate differences in program from differences in district characteristics, I purposefully selected districts that were closely matched across a series of descriptive variables. I additionally wanted districts that had adopted Q Comp early enough to allow me to explore both the initial design and any subsequent changes in program operation. Finally, I hoped to find districts that exhibited useful variation in the operation of their Q Comp programs in order to draw out important areas for study. To meet these requirements, I began by speaking with state-level officials both at the Minnesota Department of Education and at Education Minnesota, the state teachers union, gathering suggestions about districts that might offer unique perspectives on Q Comp design and implementation. After obtaining a set of districts that might potentially be of interest, I put together a table of descriptive characteristics for all Q Comp districts and identified groups of three districts that included at least one of the districts proposed by state officials and that were closely matched in terms of enrollment, number of schools, test scores before Q Comp adoption, the percentage of minority and free/reduced-price lunch students, and average teacher pay before Q Comp adoption. From here, I chose a single set of three districts that all adopted Q Comp early in the state process and within one year of each other.

This sampling protocol allowed my interviews to closely focus on contrasts across districts in Q Comp design and operations that could not easily be explained by structural features of the districts. The goal here was not to create a random draw, but rather to juxtapose a set of districts that would offer useful comparisons in defined areas and that could be used to generate hypotheses about relationships that might be later tested within a wider sample (Hartley, 1994).

The chosen research sites share several features (Table 4.1). Characteristic of the early wave of Q Comp adopters, these districts tend to be better-off than the typical Minnesota district, with larger percentages of per-pupil funding provided locally, fewer minorities, and fewer students receiving free-reduced price lunch. Geographically, each one is situated in the Twin Cities metro area, less than an hour's driving distance from Minneapolis, and these suburban districts have a considerably higher population density than the average district in the state.

B. Sites

The Farmington school district sits at the southern end of the Minneapolis-St. Paul metropolitan area. Its approximately 6,500 students are divided between five elementary schools, two middle, and one area high school. The district itself has grown considerably over the past years adding nearly 1,000 students between 2003 and 2010. During the years leading up to Q Comp, state per pupil funding dropped off in the district, leading to a loss of just over \$500 in real dollars per pupil by 2006 from a high of around \$8100 in 2003.

Less than 30 miles from St. Francis school district, Centennial School District serves roughly the same number of students as both St. Francis and Farmington, although a higher population density means that it is geographically smaller than the other two districts. Funding-

wise, the district also has a somewhat different profile; less industry in the district means that families rather than businesses pay somewhat higher marginal taxes to raise approximately the same level of local educational funds as in the other districts.

St. Francis school district sits directly at the center of Minnesota's crook, about the same distance north of the twin cities as Farmington is south. Like Farmington, it occupies the set of wealthy suburbs ringing Minneapolis and St. Paul, however the size of its student enrollment has stayed relatively flat during the same period that Farmington's expanded, hovering around 5,800. Funding trends in the district look similar to the other case studies, with state revenues falling by about \$500 in real terms before the introduction of Q Comp and then rising slightly.

Table 4.1: District Descriptives Before Q Comp Adoption

	Farmington	St. Francis	Centennial
K-12 Enrollment	4,790	5,043	6,858
Percent Minority Students	3.0	2.2	2.2
Percent of 5-17 Year-olds in Poverty	5.0	5.6	4.5
Percent Proficient on State Test	80	80	87
Locally Provided Per-Pupil Revenue	7,485	7,121	7,429
State Provided Per-Pupil Revenue	2,406	2,480	2,058
Year District Entered Q Comp	2006	2005	2006

Poverty percentages come from the American Community Survey. Revenue totals are from the Common Core of Data. All other figures are from the Minnesota Department of Education.

C. Data Collection and Analysis

Within districts, I selected interview subjects to include members of the Q Comp leadership and design teams, principals, teacher-leaders, and teachers who did not hold Q Comp leadership roles. I interviewed approximately ten people within each district, returning for second interviews with several subjects, and using later interviews with subjects in different district roles to confirm impressions raised by initial discussions. I conducted these interviews using semi-structured protocols that drew upon my previous research with state officials and

upon information about the district programs that was contained in the applications that the districts had submitted to MDE.

Initial interview coding took place during the data collection phase and I used later interviews to check initial understandings and to further develop theories (Miles & Huberman, 1994). I used a process of iterative coding, beginning with broad, descriptive categories separated by district that focused on the origins, operations, and attitudes toward each district's program design (Corbin & Strauss, 2008). As the process continued, my codes became increasingly abstract and I began to make comparisons and develop theories across case studies. I subsequently compared the grounded theories developed in this phase of analysis with various schema developed in previous studies of district capacity. During this period, I additionally began to explore my theories with key subjects' in the study, and I made use of their responses in the analysis (Eisenhart & Howe, 1992; Miles & Huberman, 1994).

IV. FINDINGS

Across my case studies, the reader will find a series of matching components – performance-based pay, teacher observation and coaching, and teacher leadership – that have been drawn together in what appear to be somewhat idiosyncratic ways. I will argue however that there are several discernible patterns of design and implementation that fit within a broader set of theories around district and state capacity.

In the following pages, I begin by describing the legislation that created Q Comp and the implementation of the program at the state level and then describe interactions between the districts and the Minnesota Department of Education as a way of beginning to explain patterns of implementation. In particular, I argue that the programmatic guidance these districts received

from the state provided the context for a design process that was highly specified in certain ways while entirely open in others. As might be expected, districts made opportunistic use of the program, with individual personalities and politics at the district-level playing major roles in adapting the state requirements to fit particular needs. But districts had the most leeway to exercise this type of influence in the realm of teacher development strategy where the state took a backseat role. This arrangement is consistent with what might be expected from a state program where the department of education had considerable experience for enforcing compliance around structural guidelines but little capacity or knowledge around instructional guidance.

As a result of these forces, program heterogeneity at the district level depended greatly on the beliefs and ideas about teacher development that developed at the district level. In the body of this section, I explore the different models of teacher learning that developed in these three Q Comp districts and the impacts that these had on the workings of the program in different districts. I argue that ideas about and strategies for teacher learning represent an important element of district capacity to consider in order to understand the design and implementation of teacher improvement systems.

A. The State Program – Q Comp in Minnesota

Q Comp took shape as a state program in 2005 when the Minnesota legislature came together around in a hurried special summer session to pass an omnibus education bill that had been delayed by debate during the previous sitting. The passage of the bill followed a long period of discussion across the state about pay-for-performance and whether or not Minnesota should directly compensate teachers for student gains, yet the law as it was finally passed took a

broader stance around teacher improvement that is strikingly similar to recent state action around the federal Race to the Top. In particular, the strategy of requiring districts to adopt a series of linked human capital components but allowing the districts to determine the final shape of the plan appears to be the standard model (Baeder, 2011).

The bill asked districts to pledge to undertake a unified program design within state guidelines and in turn gave the Minnesota Department of Education the power to increase funding by \$260 per pupil (\$190 in state grant and \$70 in partially equalized levy) for approved districts. The state's guidelines for the program, modeled in some respects after the Teacher Advancement Program, proposed that districts modify not only the way they paid teachers but also the teacher leadership opportunities they made available and the types of support that they provided for teachers around raising student achievement. Importantly, the legislation suggested that teacher incentives and teacher learning opportunities needed to be closely interlinked, with teacher professional development directly supporting the changes to the salary schedule. While the pay reform aspects of the bill garnered the most headlines, the teacher development piece was given equal or greater attention by legislators, with all but one paragraph of the final legislation devoted to the teacher support structures, including career ladder positions and professional development requirements.

In particular, districts were asked to submit plans that included:

- Salary schedule reform ensuring that at least 60% of any teacher pay increases are based on teacher performance. Teacher performance benchmarks can include school wide achievement gains, measures of student achievement at the teacher level, and/or teacher evaluation results.

- Career advancement opportunities for teachers that allow teachers to retain primary roles in student instruction.
- Site-based professional development, led by master or mentor teachers, and taking place during the school day.
- An educational improvement plan that elaborates on data collection mechanisms, links the above elements of the Q Comp program to specific school performance goals and ties these elements together into a “rigorous professional development system.”
- Joint agreement on the plan from both the local superintendent and the “exclusive bargaining representative of the teachers” (e.g. the local union leader in the case of traditional districts).

In previous work that investigated the ways that Q Comp was managed and implemented at the state level (Schwartz, 2012), I argued that limited state capacity to provide genuine instructional support meant that local districts had substantial autonomy over important aspects of the Q Comp program. While MDE set firm guidelines around application contents and program structure that clarified much of the ambiguity in the legislation, it handed off the decisions about the content of professional development to individual districts. This strategy, which left Q Comp teacher-leaders, local union representatives, and district administration to determine how to make professional development relevant, informative, and useful, opened the door to important variation in program implementation. Yet these patterns cannot be fully understood from the state level. How did districts experience the state’s guidance over the Q Comp program? And what did the patterns of state guidance mean for district program design? The three districts in my study help to flesh out this picture.

B. State-Local Interaction – Patterns of Guidance and Control

Much of the literature on district implementation implies that districts will exercise their autonomy to control program structure whenever it is in their immediate interest to do so. Thus, policymakers need to craft the appropriate blend of rewards, threats, and sanctions in order to raise the probability of compliance (Firestone & Corbett, 1988). However, several researchers have pointed out that local autonomy can also be affected by the degree of specification around particular policy demands; policies that are clearly defined and supported by program leadership will have greater influence than those that are more ambiguous (Cohen & Ball, 1999; Cohen & Moffitt, 2009). If this is the case, district take-up of state teacher improvement systems will be structured by the blend of politics and governance that shape these programs at the state level.

The findings from my study support this conclusion while adding detail about the ways that capabilities at both the state and district levels can influence the shape of implementation. I find considerable evidence of local opportunism and program adaptation tempered by the particular pattern of state action. On the one hand, members of each district that I worked with described in detail both what they had to gain from the program. Part of what made these districts willing to join was that they felt that they had considerable freedom with respect to the design of their district programs and thus would not have to transform themselves in order to implement the state's guidelines. But the story is also more complicated. The districts I studied generally acceded to the state's demands around program structure, which were quite clear and specific though not backed by strong enforcement powers. At the same time, these districts found themselves with little guidance or direction about teacher development within the Q Comp program, and they thus developed their own quite different understandings around this aspect of the program.

In each district, I was told by those involved with the initial program design that they viewed Q Comp as a financial opportunity that the state had made possible, but that they took advantage of this opportunity on their own terms rather than those set by the state.

In St. Francis, for instance, program leaders recall joining Q Comp not because they necessarily endorsed the changes demanded by the state program but because they felt that the state program could provide extra resources to support the work that they were already engaged in. The program's primary designer, John Dougherty³ recalls:

Our [professional development work] was going well. All of a sudden, the Q Comp program came up. We said, what a great fit. And we could use the extra cash... I think the state passed a law that they thought was going to be a statewide TAP program [Teacher Advancement Program]. The districts didn't treat it that way at all. In a very short period of time, the [state] commissioner began to brag about the fact that this was locally designed. I think they saw almost immediately the strengths of getting this designed around local needs.

Similarly, in Farmington, Mark Sanders, the primary author of the application and subsequent president of the local teacher's union said: "The bigger factor – it truly turns into money. If you can get extra compensation for people within a school district, extra money for them to implement different programs, extra money to pay people, you'd be silly not to go after that."

Sanders went on to observe that going after this money did not necessarily mean that the district adopted the state's vision for the program, which he saw as focused on merit pay.

³ All names are pseudonyms.

I was personally completely opposed to [merit pay], but we were given the ability with our district leadership to put something meaningful together. If we could structure the program in a way that money can go into teachers' pockets for doing something that's going to improve their classroom, we're all for it. So then we had to come up with language to satisfy our [union] members and the statute and MDE. So essentially, what we put together is a bunch of language that if an outsider reads it, you won't be able to make head or tail of it.

Reading this quotation alone, it would seem that Q Comp's operations in Farmington would turn out to bear little relationship to any sort of standardized state program. Yet the rest of my conversation with Sanders as well as my interviews with Farmington teachers and administrators gave a different impression. Each year, all of Farmington's schools dutifully send goal statements to the Minnesota Department of Education for approval, and teachers submit professional development plans that look similar to those across Q Comp districts. Teacher-leaders paid through the Q Comp program complete regular mentoring and peer observation sessions around these professional development plans. At the end of the year, teachers receive bonuses based on both their participation in these processes and various student outcomes at the classroom and school levels. Many of these operations echo what is taking place in Q Comp districts across Minnesota. Where then have districts had the greatest influence on Q Comp programming and where have they tended to look to the state for direction?

State guidelines for Q Comp have tended to be strongest in areas that invited clear reporting and compliance measures (see Schwartz, 2012, for more on the state's role). This is

one reason that most district salary schedules look quite similar after the adoption of Q Comp, since this is an area that the state has tended to focus on in Q Comp applications. At the same time, it is notable that despite the common belief that Q Comp represented the creation of a pay-for-test-score system, the state has actually been fairly restrained in its push for strict performance pay. It perhaps makes more sense to describe the salary schedules that the state required of Q Comp districts as modifications to existing schedules that offer teachers small bonuses for school wide performance, as well as payment for participation in various forms of teacher observation and classroom goal-setting. Most Q Comp districts have consistently agreed to this shift.

The state also plays a considerable role in defining the performance goal-setting process, requiring that each school set defined yearly testing goals in either mathematics or reading and secure state approval that these goals are both clearly measurable and sufficiently ambitious given the previous year's scores. Other processes that have garnered considerable state attention include the structure of the teacher observation process, the posting of openings for teacher leadership selection, and the reporting from teacher PLC meetings. In each of these cases, district personnel described the ways that they modified their original plans for the program to reflect state demands.

As a member of St. Francis' Q Comp team recalls: "The state wanted a much more formal observation process than we had originally designed. We looked at the money and said, okay, for 1.5 million dollars, we'll do this. And to be honest, I think it strengthened our program to have a very formal observation process. I think it actually turned out to be a much better program." In Centennial, district personnel found themselves repeatedly frustrated by what they found to be an opaque school goal-setting process. "There have been some hurdles there where

we will submit our goal, and although we are very open to feedback and suggestions, a lot of times we'll submit a goal one year and it's kicked back and said that's unacceptable and we'll rewrite it and that's unacceptable and then the next year we'll write it and it's fine," said a Centennial principal. At the same time, the district has willingly submitted to and learned to work within the state process. Indeed, when I spoke with members of the district, one school had just learned that its teachers would not receive the portion of bonuses associated with school performance because they had missed their goal by only 0.02 percentage points. Teachers and administrators all noted this cutoff as feeling somewhat arbitrary and unfair, but several spoke positively of the process and observed that the cutoff was the necessary and acceptable result of the greater system.

Yet where certain parts of Q Comp have been quite clearly defined by state intervention, others were left far more open. In Centennial district, for example, officials note that it has been largely left to the district and to teacher-leaders in individual schools to determine how Q Comp's professional development components will connect to the rest of the program. Since the district had placed a large emphasis on professional learning communities before Q Comp, these PLCs became one of the mainstays of the program. Whether or not these were meaningful was the responsibility of the district. One teacher-leader explains, "The districts are charged with training everybody – how to lead a PLC group and how to be a member. The state has video courses, but there's not a lot of hands-on training for sure." After thinking this over, the teacher went on to suggest what was missing from the current model:

I think MDE should be providing training for our learning community leaders who are coming in and saying, this quarter...like in the fall we're going to work

on coaching. And then in the spring, we might talk about how to get PLC members invested in the observation process and bringing that information back into their classroom and making a change in the classroom. How do you go about doing that? Instead of just telling them do it. I do the best I can with trying to get resources for my LCLs, but I just don't have expertise in coaching. That's not my particular part I do well in. It would be nice to have more help and training for the people you are asking to lead.

Teacher-leaders in Farmington offer a similar view of a program that has been very defined in certain respects but that leaves something of a vacuum around the question of what teacher professional development should actually consist of. Even in areas where the state provided some guidance, Farmington leaders say that this guidance gave them little help in determining the program focus. As one teacher leader explains: "The state would say, we provided a rubric and gave standards. And to some extent, they did. But once the application is approved, it takes on a life of its own in the school district."

These patterns of interaction around Q Comp between the state and its districts have created an interesting program dynamic. The state's intervention has been most defined and specific on the side of program process and incentives including school-level goal setting and teacher-level pay, and this means that in many respects district Q Comp structures look remarkably similar for a program that seems quite undefined in its original legislative form. At the same time, the state has taken a much less active stance on the teacher learning side of the program. This outcome maps a common pattern where highly constrained state agencies tend to be strongest on managerial requirements and weakest on guidance for implementation (Cohen &

Moffitt, 2009). As a result, districts in Minnesota have been left to determine for themselves how teacher learning will be designed and whether it will encourage the kinds of improvement that the program is meant to sustain. In the following sections, I describe program operations in each individual district. I begin by briefly summarizing program structure, highlighting the similarities across districts before transitioning into the crucial substantive differences that I found in program type and the implications of these designs for the development of cognitive capital within the district.

C. Q Comp in Farmington

i. Program Structure

The Farmington School District joined Q Comp in the 2006-2007 school year, a decision driven jointly by district and union representatives who looked to the program to provide a supplementary source of teacher funding while opening up new opportunities for teacher leadership where few such positions had previously existed.

With respect to its alternative compensation plan, Farmington looks quite similar to many other Q Comp districts. The pay incentives that accompany the program add a total of \$2,000 on top of the teacher's base salary. These incentives are divided into three parts so that teachers receive one portion of the bonus for the submission and completion of professional development plans (PDPs), one portion for three successful teacher-to-teacher observations, and one portion for the attainment of a building-wide goal for improving student reaching achievement. Importantly, however, with the exception of the building-wide growth goal, these extra pay components are less a function of teacher performance than they are of teacher effort levels.

Individual-level pay bonuses in Farmington do not get withheld as the result of poor quality observation sessions or PDPs but only for the failure to complete one of these elements entirely. In practice, this means that most teachers receive at least a large percentage of the extra pay as the result of taking part in the series of teacher learning opportunities that make up the district's Q Comp program.

Similarly, the set of teacher-leadership opportunities, observation procedures, and professional development options included within Farmington's Q Comp plan do not immediately distinguish themselves from other districts. Indeed, the structures listed within Farmington's Q Comp application seem to exist in nearly every other Q Comp district. Under the current system, Farmington teachers apply to a central committee to serve in roles such as mentor teacher, peer coach, and facilitator of instructional support. These leaders in turn help teachers to participate in a year-long process that includes learning communities, individual and school goal-setting, peer coaching and evaluation.

However the structural similarity between Farmington and other Q Comp programs mask a few key points that serve to distinguish this district's particular program design. As expected from the pattern of state oversight, these differences lie in the program's approach to teacher learning and the means that the program calls on for bringing new ideas and new expertise into teachers' classrooms.

Across interviews, district personnel highlighted two aspects of Farmington's Q Comp program that I did not find in my other case studies.

First, the program that the planning committee put together reflected a district characterized at the time by considerable mistrust between teachers and the administration and, correspondingly, little centralization at the district level. Unlike in many districts, nearly all

management of this system has been delegated to individual schools committees, known as Farmington Alternative Teacher Professional Pay System (FATPPS) committees or more simply as the “Fat Peas.” From here, oversight tends to be delegated even further so that most of the programming and decision-making takes place in individual meetings between teachers and instructional coaches.

Second, and closely related to the first, Farmington’s Q Comp design draws on individual teachers to provide the person-to-person guidance that is meant to stimulate improvement across the district. According to interview subjects, the program revolves around individual teachers’ professional development plans and subsequent peer coaching process. When district personnel at all levels talked about teacher improvement within Q Comp, they tended to be referring to the individual observation sessions that teachers completed regularly in each other’s classrooms. This stance toward teacher learning, which suggests that teachers who want improve their practice will find the most useful learning opportunities by observing their peers and sharing practices, represents a view of the teacher learning process that differs considerably from what I found in my other two case studies. I describe the approach in more detail in the next section.

ii. Teacher Development in Farmington – Learning from Other Teachers’ Practices

Embedded within Farmington’s Q Comp program is a particular set of ideas about instructional expertise. In this structure, instructional expertise is viewed as something that already exists within the district but that needs to be better harnessed and spread across all classrooms by way of the Q Comp program. Teacher improvement, if it occurs, will take place because teachers are drawing new knowledge from *other teachers’ practices*, and this process is meant to be driven by program elements such as peer coaching, teacher observation, and best-

practice sharing. While some of these program components exist in the Q Comp designs of my other case studies as well, the focus is quite different and these other designs emphasize other sources of expertise, specifically expertise provided by student data or external professional development.

What does the Farmington design look like on the ground? Interview subjects consistently highlighted the frequency of teacher observations and the growing comfort that teachers had with peer observers in their classroom. As one teacher observed: “Open doors – that’s our culture now. I don’t think that was our culture before this program.” But there remains a real concern among many subjects about the degree to which the design is generating substantial teacher development.

To make sense of this concern, it is helpful to begin with a description of the Q Comp implementation process in one elementary school by the school’s Q Comp coordinator, Dwayne Eriks. Initially, Eriks explained, program coordinators focused their energies on making the improvement process acceptable to teachers who did not like the idea of having regular observers in their classroom.

“It didn’t take hold right away,” Eriks explains. “It was very uncomfortable at the beginning. Nobody really likes having anyone judge them. But once people tried it, they realized it wasn’t an evaluation, it wasn’t a judgment, it was simply a coaching session.” To make the process feel less stressful, the school decided in the second year of the program to expand the definition of peer coaching so that teachers could either choose to have the designated peer coaches come to their classroom to critique their lessons or alternatively they could visit one of the coach’s classrooms to pick up a new set of teaching practices. “It takes the pressure off the teacher,” Eriks says, “and makes it more user-friendly.”

But although this modification helped to quell initial distrust of the program, it didn't solve the larger problem of how to make these coaching sessions a meaningful source of teacher improvement. Indeed, according to Eriks, the change only highlighted a broader problem concerning the content of the peer observation sessions.

A lot of teachers didn't know what to ask the peer coach to come in and observe.

That was the tricky thing. Peer coaches would come in and say what do you want me to observe? And people would say: I don't know, how many boys versus girls do I call on during the class? It was a kind of petty thing, we thought.

As a solution, Eriks and the rest of the Q Comp leadership in the building sat down to clarify the process, putting together a list of potential teaching strategies to be covered in peer observation and coaching sessions (Figure 1). The points on this list provided an easy reference list for teachers to turn to during observation sessions and to some extent they offered a solution to the problem of how to generate focus areas for peer observation. Yet as a tool for teacher improvement that is meant to work within the larger Q Comp structure, the list leaves something to be desired. The topics included – wait time, classroom management strategies, higher-level questioning skills – remain quite broad and generic, disconnected from the rest of the Q Comp program and unrelated in any direct way to student achievement goals or professional development plans. Moreover, it would require a highly skilled observer to be able to offer the kinds of feedback within these topics that would result in substantial instructional improvement. A different way of understanding these concerns is to note that the Farmington design, which suggests that teachers will each draw expertise from their peers, places the onus for improvement

entirely on individuals and it demands that those individuals in the system have the ability to accurately identify and pass on their own best practices, setting a high bar for progress within the system.

These demands help to explain the common refrain within the Farmington district that progress within the Q Comp program is highly individualized, with the program giving teachers only what they are willing and able to draw from it. According to the Director of Teaching and Learning in the district, “Part of [the Q Comp experience] depends on who the teacher is and how much they value the PDP work. The teachers that strongly value it make it a meaningful experience for them and it impacts their every day. Those are the teachers that say that they are learning about a particular staff development piece and they build that into their PDP. They certainly make it much more meaningful and impactful in their daily life.” Similarly, a Q Comp teacher-leader observes that the meaning of the program is nearly completely determined at the classroom level. “Your professional development plans and portfolios are what you put into them,” she explained, before going on to note the uses she has found for the Q Comp process in her own classroom.

Understanding why this particular program style and emphasis took hold in the district requires looking back at the district’s starting point when it applied to join Q Comp in 2006 and the set of constraints and capabilities that the district brought to the initial design process. As the district’s superintendent for teaching and learning explained, the district differed from a number of Q Comp applicants in that it had a history of mistrust and relatively little cooperation between teachers and administrators. This made the local union highly unwilling to sacrifice individual teacher control over any aspect of the program design. “Early on, it was an either or thing,” the superintendent said. “Either the union will do all of this, or they weren’t going to do it. This

happened because there was a trust breakdown between the administrators at the time that were making some decisions and some of the teachers.”

Mark Sanders, a teacher and union representative who served as one of the plan’s primary authors, tells a similar story, but he emphasizes his and the union’s drive to incorporate building-level choice and include as many teachers as possible in the process.

I had just gotten done with a lot of graduate work, and I was a big [Robert] Marzano fan. One of things he advocated was that every building is unique and different because of students, because of faculty, because of physical surroundings, etc. When we were building this thing, we wanted to give individual buildings the ability to meet the needs of their culture. For example, a building with 25 teachers who have only taught for 5 years versus a building with all experienced teachers...those will be entirely different schools. Our plan gave enough leeway and the ability to make decisions for individual buildings and for groups of team teachers to do what they think is best.

What developed out of this planning process was a program with little centralization at the district level.

Nearly all Q Comp decisions get delegated first to individual schools and then to teachers via the peer coaching process. As one member of the Farmington central office explains, potential difficulties have not been worked out in a standardized manner, so “all problem solving falls to the building itself.” In turn, the program itself tends to succeed or fail in individual

classrooms as teachers chart their own courses, making some use of the system or, in other cases ignoring it entirely.

Importantly, once this model became the basis for the program, Q Comp leaders found few reasons to revisit the design process since they had become increasingly dependent upon the program as a small but significant source of funding for teacher leadership positions and for additional salary bonuses. Importantly, nothing in the state's program design requires regular updates and modification to the program, meaning that there is little incentive and several disincentives for districts to undergo a systematic process of program development and improvement.

"[The program] hasn't been modified since the application was written, which it probably should be," says one teacher-leader. "There was some fear that perhaps it wouldn't get approved again." This fits a pattern that I found more generally in each one of my case studies. With little pressure from the state for substantial revisions and little guidance as to how these revisions might work, districts have tended to stick closely to their original Q Comp designs, further solidifying the role played by initial conditions in charting the program's course.

iii. Implications for the TIS and the development of cognitive capital

I argued above that Farmington's Q Comp program offers a set of teacher professional incentives and program structures that are relatively similar to those across the set of Q Comp districts in Minnesota but that the district's design also includes a specific set of choices about the way that instructional expertise is integrated into the broader program. What do these design choices imply for the program as a whole?

One concern is how these choices affect the coherence of the program and the likelihood that teachers are building cognitive capital around the particular types of knowledge that would lead to genuine student improvement. In Farmington, teachers work individually and in small groups on the problems of practice that they themselves identify as important to their larger professional development. However, there is little within the program design that ties these professional development experiences to anything that is taking place on a larger scale in the school or district. According to several teachers, the discussions that take place through the teacher coaching seem quite disconnected from the larger set of program components and incentives.

This observation ties to another set of concerns around the development of a teacher knowledge base. The internal learning model emphasized by Farmington constrains the scope of teacher learning to only the particular forms of instructional knowledge that exist in the district and are identified by individuals as something that is both integral to their success and can be passed on to others. Here, the example of the elementary school where teachers often failed to come up with useful topics for peer coaching sessions is quite telling. In the midst of everything else that is going on during the typical school day, it can be quite difficult for individuals to create the conditions for genuine learning even when they are provided with a ready and willing peer to offer feedback.

One Farmington teacher summarized the situation as follows: “A lot of things we do...we love having teacher-led staff development sometimes, but only on some things. Teachers like to hear certain ideas from professionals that know what they’re doing. I think it’s frustrating for teachers because it can feel like the district is trying to save money by having teacher leaders run PD. Some people do like to have those professionals come in.” Without such outside

professionals, and without a more defined means for sharing ideas beyond individual coaching sessions, the process has been quite limited. There has been little development of a collective knowledge base and that teachers often find themselves grasping for a more defined set of skills and expertise.

This model has by no means been entirely negative. Indeed, it might be crucially important to simply provide teachers with the space to determine the course of their own development, whether or not this always leads to immediate instructional progress, since it could contribute to a greater professional satisfaction and more awareness of the improvement process. As one principal said: “I truly do believe that teachers through the Q Comp are given time to be reflective practitioners. Time is the luxury and we don’t have enough of it. Through the use of Q Comp, teachers are given that time. And they’re compensated for it.” But Farmington’s program does demonstrate several constraints particular to this particular model of teacher development.

D. Q Comp in Centennial

i. Program Structure

Like Farmington, Centennial joined Q Comp in 2006-2007, during the state’s second round of Q Comp applications, a move that officials describe as primarily a funding decision that allowed the district to capitalize on the strengths of its initial teacher development system. As one administrator explained, “Our parents pay high taxes and per pupil, we get a lot less than even our adjacent districts. We thought that this was a way to help teachers get a raise without having to tax families anymore. We used it as a supplement to our pay scale and added additional work for the supplement. The way that it’s used, it’s really not considered a raise but it was one

of the ways we thought that we could get some money into teachers' pockets." Perhaps reflective of this drive to find new sources of district funding, Centennial's process for joining Q Comp was less union-driven than the other two districts. Although the program received the required union leadership, the application process was led by the central administration and the program leadership is almost entirely separate from union leadership, a distinct difference from my other two case studies.

On paper, Centennial's Q Comp program, known within the district as the Centennial Alternative Compensation System (CACS) looks quite similar to the Farmington program. In Centennial, as in Farmington, teachers receive the majority of their bonus pay for taking part in observation and reflection sessions and attaining individual professional development goals with small additional bonuses awarded for the attainment of student learning goals at the school level. As a whole, the bonus amount totals \$2,000, and with the exception of the \$100 designated as the school achievement reward, these bonuses rarely get withheld for anything except non-participation.

Likewise, the structural elements in Centennial's program resemble those in other districts, although it is worth noting that the Centennial Q Comp program does not create as many teacher leadership and coaching positions as Farmington's. Nevertheless, there are many similarities. Centennial's program designates teachers to serve as peer coaches and instructional leaders, and creates a process that includes professional growth plans at the individual level as well as school-wide achievement goals. Teachers meet regularly within professional learning communities and they interact with program leaders around a series of focus areas.

Again, though, the structural aspects of the program hide central differences between Centennial and my other case studies in how instructional expertise gets defined and integrated

into the program. In Centennial's case, the key point involves the use of professional learning communities (PLCs) and the directives that these PLCs have been given to make use of student data to inform teachers' instruction. Where Farmington's program focused almost entirely on the peer coaching and teacher goal-setting process, Centennial's program creates an additional push around interpreting student data within the bounds of the PLC. This approach has substantially different implications for teacher knowledge building and for the Q Comp program more generally.

ii. Teacher Development in Centennial – Learning from Student Data

Like Farmington, Centennial's Q Comp design is inwardly focused, relying on district personnel to provide the knowledge that will lead to better teaching. Centennial's program includes many of the peer-focused learning elements that were present in Farmington. Several participants in the program spoke about the importance of opening classroom doors to create new learning opportunities. One teacher captured this view particularly clearly, saying: "I've always said, we don't need to go to workshops, we've got all the people we need right here if we're willing to let each other watch each other." Yet in Centennial, this internal emphasis was accompanied by a different view of where teachers will access new ideas. According to participants, Centennial's Q Comp program aims to change teachers' instruction by raising their *awareness of student data* rather than by looking to their peers for new techniques. In this conception, data itself and its disaggregation provides the new forms of knowledge about practice that might allow teachers to begin the process of instructional improvement.

The data-driven approach to teacher learning can be found most clearly in the professional learning communities that are a central piece of Centennial's Q Comp design. While

PLC groups are a component of many Q Comp plans, Centennial teachers and administrators spoke repeatedly about the primary importance of these PLCs as a means for teachers to review successes and failures around individual elements of the site goals as captured by the district's formative benchmark tests supplied by the Northwest Evaluation Association. PLC leaders thus play a major role in helping other teachers to make sense of the information. "[PLC leaders] blend the data, mine it, and go back and the PLC facilitator leads their PLC to say here is where we are at," says an elementary principal. A teacher and previous PLC leader described the progression from district to school to PLC goal-setting and data analysis like this:

Our district goal is centered around reading, so that means our site goal is centered around reading, and that means our PLC work is centered on these goals. We identify our lowest quartile and figure out ways to incorporate reading in our [curriculum] areas and how we can enhance reading for these groups in particular.

The teacher went on to explain that this data analysis is then meant to translate back to inform the greater cycle of peer-to-peer observation:

The most valuable thing is the ability to observe your peers and look at how they are doing things and then support them in the things that they want to do with a kid or with a group by having this hard cold data. As you know from being a teacher, when you're delivering the instruction, you don't really have the same kind of perspective on what's happening with the kids.

The view designed into the Centennial Q Comp program that teacher development can be driven at least in part by student data is one that has propelled the rapid spread of the formative and benchmark assessment industry across the country (Burch, 2006; Massell, 2001), and several principals and teacher-leaders in Centennial spoke positively about the process for integrating testing data into teacher professional development. “What we do is get data teams together, to look at the data, the different strands,” explained one principal. “Then [since the school is focused on math improvement] we’ll send math teachers within the various PLCs to talk about what are doable goals.” Another principal noted that the data emphasis in Centennial was not new but that Q Comp had raised expectations for teachers’ role in data analysis by creating paid positions for teacher-leaders interpret student data. “The primary change [in the district] because of CACS is now the compensated positions of leadership within our building. They blend the data, mine it, and go back and the PLC facilitator leads their PLC to say here is where we are at,” the principal said.

At the same time, Centennial interview subjects described several challenges that they encountered in bringing their particular Q Comp version into existence, which speak to the larger challenges of this model of development within a teacher improvement system. The central issue that district personnel brought up concerned teachers’ capabilities to make meaningful use of student data and to transform this data into professional knowledge or instructional practice. In many ways, this difficulty resembles the concerns in Farmington around getting teachers to correctly identify and make use of best practices, and both are typical of inwardly focused programs of teacher development; improvement in these programs depends entirely on the current capabilities of the people who are being asked to improve.

One teacher, for instance, pointed out that the program's structure required that teachers have the ability to determine how to respond to the new data demands. "It is all based on our own research. We read articles. We talk to people in other buildings. But we need to figure out how we can get more information. I don't know where other PLCs get their stuff." Several others noted the insufficiency of the current training system and the demands it placed on individual teachers. For instance, one teacher who had served as a PLC leader noted that many of the group leaders felt insufficiently secure in their own data skills to lead others in the same task. "We could use more training for people. In how PLCs work. Even just educating about what a PLC actually is and how we are using PLCs. Is there a better way to be doing the PLCs in our district?" Another district Q Comp coordinator tied the lack of training to the ways that the district had decided to distribute its Q Comp resources. "One thing that keeps coming back is that we don't have enough training for our LCLs," she said. "That is a matter of time and money—I do 4 meetings [at the beginning of the year], but it's not enough for them to become educational leaders."

The repeated calls for outside training from people who hold different positions in the Centennial district suggest both some of the difficulties with Centennial's model of Q Comp and also some possible ways that the program of teacher development might be strengthened. As with Farmington, I draw out these implications in the next subsection. But first it is important to ask why this particular model of Q Comp developed in Centennial, especially since the process of data-based learning is a popular concept that has been used in all three of my case study districts to justify the use of quarterly or benchmark assessments, and it is also one that receives strong support from state officials.

As in Farmington, the answer seems to be partly a function of the personalities who influenced the original application and partly the result of the way things worked in the district before the state program came along.

The influence of previous district policies appears to have been especially strong in Centennial. Much more than in the other two cases, Centennial interview subjects stressed the fact that their Q Comp application was driven by the financial needs of a district with a small industrial tax base and an already high individual tax load rather than direct ideas about reform. To join the program, Centennial's Q Comp leadership made the necessary changes with respect to salary and teacher evaluation, but where teacher learning was concerned, they focused on what already existed in the district. The district had previously set up regularly meeting professional learning communities and already emphasized the need for using data to set so-called SMART goals (specific, measurable, attainable, realistic, and timely). These elements were simply transferred over to become part of the Q Comp program. "A lot of the pieces, we were already doing. So it really fit in," says one administrator. Another who transferred to Centennial district from another Minnesota district just around the time that Q Comp began detailed her immediate impressions of the teaching force. "They have been doing legitimate PLCs and smart goals for going on 12-15 years. These folks understood how you write a smart goal, collect data, change your teaching practices using the data," she said. In these circumstances, "to adapt the Q Comp to make it ours was just natural if we could be paid more, for things we were already doing." These sentiments echo several other interview subjects. They also suggest both the benefits and disadvantages of such a course of action. On the one hand, the adoption of Q Comp in Centennial looks less like a break from the past than a set of tweaks to already existing programs. On the other, the district has been able to call upon considerable experience with many of these

elements and, according to several administrators and teacher-leaders, Centennial's program benefitted from both the previously existing knowledge and the rapid buy-in that this made possible.

iii. Implications for the TIS and the development of cognitive capital

While Centennial shares many elements of its Q Comp structure with my other case studies, the district has adopted its own data-centric approach to teacher development, emphasizing the need to make instructional decisions based on disaggregated student data. Again, these design choices present a particular set of dilemmas with respect to the broader teacher improvement system and the development of teacher cognitive capital.

The benefit of this approach is that district personnel are focused on testing information tailored directly to the needs of teachers and students in each school. Moreover, the design creates a system that is far more coherent in aligning teacher incentives and professional development than the Farmington version since the data focus is highlighted across every aspect of the program.

At the same time, Centennial teachers still expressed some doubt about the strength of the information they were receiving from the various tests used within the district, and several teachers expressed frustration at the fact that test results sometimes seemed quite divorced from what was taking place in classrooms. This frustration was particularly pointed in the school that had recently learned that its teachers would not receive the school achievement bonus payment because the school had missed its testing target by 2/100ths of a point. "We just found out," a building coordinator explained. "It's a couple hundred dollars to each staff member, that in itself is not such a big deal, but it's just... Every year it will be harder and harder. What I have to tell

my staff is, you know what, we're improving instruction, are kids are getting better, we can't worry about whether we achieve that 100 percent or not." These issues speak to the disconnect that can exist, regardless of programs like Q Comp, between the large-scale assessments increasingly used to measure student progress and the day-to-day work of teaching. While programs like Q Comp that push for a greater focus on student improvement outcomes can increase the drive to pay attention to the outcomes on particular tests, they do not address the doubts that many teachers hold about the ability of the tests to measure the quality of their teaching.

In many ways, the Centennial Q Comp design seems more likely to address the problem of building a collective knowledge base than the Farmington plan because the types of group analysis that occur place a greater emphasis on collective improvement. Through jointly managed student analysis, the Centennial design creates a standardized push for similar sorts of improvements across classrooms and schools. However teachers' calls for more external training as they work to adapt the data-based learning into their own instructional practice are telling. As the Centennial system is currently designed, it requires that teachers have the initial skill base to allow them to make good use of student data, and many interview subjects expressed hesitation that this was the case. Even more concerning, Centennial district personnel had substantial experience working in PLC groups and working with student data long before the district joined Q Comp. For districts that have had a weaker emphasis on this kind of teacher development process, the transition is likely to be far more difficult.

What we see in Centennial is another version of the Farmington dilemma. Teacher development is meant to take place while relying solely on resources and ideas about teaching that already exist within the district. Teachers offer some evidence that they find this process

useful but they also note that the system has done little to resolve many of the greater uncertainties they feel about how to develop their craft. This dilemma provides a useful backdrop to the alternate model found in the St. Francis school district.

E. Q Comp in St. Francis

i. Program Structure

St. Francis offers an example of a Q Comp design that is strikingly different from my other two case studies. The district was one of an initial group of 6 traditional school districts to join Q Comp in 2005-2006, the first year that the program was offered. This decision to enter the program was driven by a group of influential union representatives who in previous years played a critical role in charting the district's course around teacher development. Indeed, many of the elements that came to define St. Francis' Q Comp design had taken shape several years earlier during a set of district-union negotiations over teacher absences from professional development events. These negotiations created the "Teacher Academy," a set of professional development courses funded by the district and taught by local teachers after several weeks of off-site training through the American Federation of Teachers' Education Research and Dissemination (ER&D) Program.

As the Teacher Academy grew, John Dougherty, the local union's lead negotiator and the driving force behind the Academy took it upon himself to push for its integration into the state's Q Comp program. He explains that the district had explored the possibility of joining the state's initial pilot program several years earlier, and while this had proved difficult as the teacher academy initiative was just getting off the ground, it left the district well positioned to join Q

Comp as soon as it launched. Most importantly, the academy's focus on teacher leadership and its negotiated place in the district with teacher academy course completion substituting into the traditional steps and lanes system made it a natural fit for Q Comp. "My deal during that whole period of time was to say look, it really looks like this thing [e.g. Q Comp] is coming like a train going down the tracks," Dougherty says. "I told people you had to figure out whether you want to steer it or jump on as it goes by or get run over."

The district-level Q Comp program that emerged on the other end of that process continued to rely on St. Francis' initial professional development design anchored by the AFT's ER&D course, but it also incorporated Q Comp's observation and incentive pay elements into what teachers still referred to as the Teacher Academy. Thus, St. Francis includes several of the teacher observation and pay elements that are common among Q Comp programs, but the central role of the Teacher Academy has created a unique program.

On the pay side, St. Francis Q Comp teachers enter into a salary schedule that bases pay partly on the results of teacher evaluation and partly on the results of individual goal-setting procedures. Like in the other districts that I studied, very few teachers receive the failing designations that could result in the loss of pay and, since teachers have the option of repeating a failed process, almost no one actually is denied these pay raises. However, St. Francis' salary schedule also represents something of an exception among Q Comp districts. Where most districts offer teachers pay bonuses for reaching various goals, St. Francis' plan modifies only base pay, a decision that reduces the year-to-year uncertainty that teachers face and further emphasizes the importance of professional development inputs. Additionally, while St. Francis schools are required as part of the state program to set student achievement goals that result in a school-level reward, the reward for meeting these goals is not parceled out among individuals

but rather given as a payoff to the site as a whole, to be used for building needs like extra computer equipment. These decisions fit with the overall emphasis in St. Francis on Q Comp as centrally a professional development program with supporting incentives rather than vice versa.

To some degree, this professional development looks structurally similar to what I found in Farmington and Centennial, and to what MDE pushed for at the state level. Teachers in St. Francis put together individual goals and work with teams of teacher-leaders who coordinate regular observations and feedback on these goals. Meanwhile, larger groups of teachers meet regularly in learning communities to assess progress toward school wide goals.

But what makes this structure operate entirely differently than in the other districts is the presence of the Teacher Academy and the AFT's ER&D program. Together, these initiatives provide the content that drives the other elements of the Q Comp process. In the expanded version of the Teacher Academy that accompanied the adoption of Q Comp, teachers are required and school administrators are encouraged to attend 32 compensated hours of courses each year. The courses are taught with curriculum provided mostly by the AFT by district teachers who attend multiple weeks of AFT training. These courses then form the basis for the rest of the district's Q Comp program.

While enrolled in Teacher Academy courses, each individual teacher meets with his or her performance review team (PRT), which is led by the course instructor, along with several other trained observers. During this meeting, the group agrees on classroom-level goals derived from the particular course curriculum. As the year progresses, the teacher then collects evidence on progress toward these goals and is regularly observed by the PRT. Eventually, the PRT offers the teacher a summative rating on the process ranging from incomplete to fully engaged that helps to determine the teacher's eligibility for future pay raises. In the following section, I argue

that the presence of both the external AFT curriculum and the Teacher Academy within St. Francis' Q Comp design created a particular style of program that allowed different forms of teacher learning and thus a considerably different improvement system than I found in my other case studies.

ii. Teacher Development in St. Francis – Importing External Knowledge

St. Francis' Q Comp model differs from the other case studies because the program has been designed to import instructional expertise into the district and to transmit this expertise by way of the various program structures such as observation/evaluation protocols and teacher leadership roles. Teacher improvement in this model depends as much on an external vision of what good teaching might look like as it does upon the communications between teachers within the district. This version of Q Comp, which is built around an *external source of expertise*, offers a novel view of the way that a teacher improvement system might operate at the district level.

As noted, the element that makes St. Francis' program unique is the Teacher Academy, the union-led organization that serves as the basis for the entire Q Comp program design. Paid for by a combination of Q Comp and district professional development funds, the Teacher Academy is led by classroom teachers who apply to the central Q Comp committee for instructional leadership positions. Yet while the Academy is staffed internally, the training for teachers who teach there and the content that gets taught is coordinated through an AFT program known as Education Research and Dissemination (ER&D). This means that there is a direct source of supplementary ideas about instruction that is directly integrated into the larger Q Comp program.

St. Francis teacher and Q Comp instructional specialist William Gibbs, who has taught math in the school district for over 30 years, provides a useful perspective on the program and in particular on the connections between Academy courses and the Q Comp evaluative process. As an instructor in the Teacher Academy, Gibbs provides his fellow teachers with 32 professional development hours of direct instruction, but he additionally serves as a member of a series of performance review teams to observe and assess teachers on their implementation of his lessons. Gibbs says that the follow-through between course content, teacher observation, and teacher incentives serves as the central innovation of the model:

[As your instructional specialist], I would sit down with you and we'd decide what you will implement. You'd say the things that I will be looking at. You'd teach your lesson and then you and I would talk afterward about what worked and what didn't work. The observation becomes very specifically related to the PD that you are engaged in during the year.

According to Gibbs, his many years of classroom experience provide an important backdrop to this process, giving him credibility as an instructor. However Gibbs downplays the impact of his classroom experience on the content that he teaches. Instead, he credits ER&D for making the district's Q Comp design work by providing high quality professional development content and training for the district's instructional leaders. "One of the things that has been the cornerstone of why we have had so much success is the training that we go through," he says. "Before we put anybody in front of a group of our teachers, we put them through some very rigorous training. I was out there [in Baltimore] for 5 days and then a 3-day follow-up prior to

ever setting foot in front of my peers.” Another teacher-trainer offered similar views on the way that the partnership with the AFT made the greater program possible. “In [ER&D], we found a vehicle to deliver high quality professional development that was research based... And it was successful and teachers bought into it and the school district bought into it.”

Through the district’s partnership with the AFT, St. Francis’ Q Comp program emphasizes a particular set of ideas about teacher learning where the additional skills and knowledge that teachers are meant to gain are seen as something additional and supplemental to what already goes on in St. Francis. These skills and knowledge are then meant to be incorporated into individual classrooms through a combination of teacher training, incentives, and evaluation.

Importantly, this setup creates a process that is quite transparent about what teachers are meant to learn and how improvement is supposed to progress. As one principal explains: “It isn’t very often that you have a district that says: here’s what we want you to improve on as a teacher, and we are going to observe you doing that, and we are going to measure how it improves your students’ performance.” Moreover, St. Francis’ strategy of training its own teachers to deliver external content retains some of the benefits of local control even as it draws in outside expertise.

First, from a financial standpoint, this model keeps much of the Q Comp funding in teachers’ pockets, an outcome that the union placed high on its priority list. Second, the system lets teachers tailor the AFT content to their knowledge of the district and appears to create a collaborative environment around this content. One teacher describes the move toward a “common instructional language.” As he describes, “The [ER&D content] gave names to a variety of things that we intuitively knew, and then we had a label and could talk about them –

like ‘with-it-ness.’ It created a lot of conversations among teachers. You could actually talk about what worked, why it worked, and what didn’t work.”

At the same time, the success of St. Francis’ approach depends in large part upon the quality of the content provided by the external partner. And since the external partner is providing course content that can be used across a variety of districts, the actual curriculum must be quite generic. As a result, teachers in St. Francis lose out on some opportunities to work on the particular problems of practice that they encounter as a result of their particular instructional framework or subject area. One teacher notes, for example, that the move to Q Comp based professional development has taken away the chance to develop expertise in the content that he teaches.

As a social studies teacher, there are now zero dollars available to me to receive any development related to my daily teaching. I can learn that I should wait for two minutes after calling on someone, to praise kids for a job well done, to encourage kids, the value of communicating with adults. But there are zero opportunities to have staff development related to social studies and more specifically to world geography.

Others complain that the Teacher Academy course content is not entirely useless but simply not sufficiently meaningful to justify the entire Q Comp apparatus. As one teacher observes: “An academy class is not cutting edge stuff. It’s mostly seat time. I sit in these classes and the 32 hours really could be covered in 4 hours.” To some extent, this is the opposite problem from the one in the Farmington school district. By coordinating with an outside partner,

St. Francis has been able to bring in new expertise around teacher improvement and to create a process that offers clear ideas about best practices, yet this expertise can feel entirely generic to the teachers in the district. Thus, some teachers have strong doubts that the new best practices imported from the AFT will actually serve to boost student achievement in the areas where teachers need the most help.

I expand more on these themes in the following section, but it is useful first to consider the origins of this model of Q Comp in St. Francis. Here again, past practice filled the vacuum left by the lack of state guidance on how to integrate teacher learning into the Q Comp program.

According to the subjects that I spoke with in the district, two key pieces made this particular style of program possible. The first was a measured process of program development over a long period of time. While Q Comp only began in 2005, the district had been experimenting with the Teacher Academy model of professional development since 2001, and discussions around these issues began even earlier. As one member of the initial design team explains:

It took us 10 or 11 years to become an overnight success... We had so many steps that led up to the big bang. Everybody was kept kind of abreast of what was taking place in increments as we were going. It was never a huge swing one way or the other... A lot of the districts want to do our plan – ours is unique in that we had steps going on already. This just seemed like another step. You've already been doing these things, now we're going to do this. You could see where it was heading. Somebody was laying it out.

This process meant several things for the Q Comp design team. Not only did the team already have substantial buy-in from constituencies that included administrators, teachers, and union leadership, but the district already had relationship with the AFT as a partner in teacher development. Among other things, this partnership made a difference from a financial standpoint, since the district was already paying for AFT services and this meant that the partnership did not seem like it would divert Q Comp money from being used toward teacher salaries.

The second factor that shaped Q Comp in St. Francis was dynamic leadership by a figure who crossed various district factions and who possessed a particular vision for the program. Many of the subjects that I spoke with gave full credit for the program to John Dougherty, who served as the union negotiator who first brought the Teacher Academy into existence and who developed a close relationship to the superintendent that made the subsequent Q Comp design possible. Not only did Dougherty draft the plan, but he has remained intimately involved or the seven years following the program's adoption, eventually leaving his post as a teacher to serve as a full-time program coordinator. Importantly, Dougherty's influence, which I was repeatedly told cannot be underestimated, makes the model difficult to replicate in other districts where such effective teacher leadership does not already exist.

iii. Implications for the TIS and the development of cognitive capital

Where Farmington and Centennial's Q Comp designs turned the spotlight inward to find a source of instructional expertise that might support the other elements of the Q Comp program, St. Francis looked outside the district. What makes this model particularly interesting is that the AFT's ER&D program in St. Francis to some degree represents a substitute for what the state

could not or would not provide. In other districts, I repeatedly heard district personnel complaining about the lack of state support and training for teachers' learning goals. I never heard the same complaint in St. Francis since the district's program design had found an alternate provider and thus did not rely upon the state for such things. The addition of an external influence meant a different set of uses for the same sorts of program structures that I found in my other case studies. What are the implications of this design for St. Francis' Q Comp program as a teacher improvement system?

Surprisingly, St. Francis' integration of an external provider into Q Comp did not appear to come at the expense of a less aligned or coherent program. Most of my interview subjects described a professional development curriculum that had been integrated quite remarkably into the greater Q Comp program by tying teacher evaluation, coaching, and rewards directly to course content. There was nothing in this case study to suggest that the integration between external expertise and internal Q Comp design is by any means a given, yet St. Francis' example also demonstrates that it is a possibility.

Moreover, unlike in Farmington where teachers seemed to be interacting with each other as individuals and not building toward a group knowledge base, St. Francis has identified a program model that emphasizes the importance of group knowledge. One of the most frequent descriptions that I heard about Q Comp in St. Francis was that it had created a shared language and a shared set of understandings about instruction. Because teachers took part in common lessons about instructional content, they could speak to each other across this content and make use of other teachers' knowledge in a way that appeared to happen far less in the other districts.

These points provide a strong series of recommendations in favor of this particular Q Comp design model, aligning with a broader set of theories about the consequential role that

non-governmental agencies tend to play in the U.S. system due to weak governmental capacity (Rowan, 2002; Foley, 2010). They also offer some suggestion as to how the Q Comp program overall might be strengthened, either through the state trying to move into the role played by the AFT in St. Francis or alternatively through a set of more standardized external partnerships supported by the state. An interesting side note here is that the current setup of Q Comp at the state level likely dissuades districts from setting up the kind of external partnerships that I found in St. Francis, since districts tend to view the funding as a source of supplementary teacher pay and thus would prefer to use all the extra funds internally rather than route any of this money outside the district.

At the same time, St. Francis' design also demonstrates the limits of this model in a broader system that makes it difficult for external providers to focus their offerings on teacher development. In order to serve as the basis for PD that can be used across subjects and grades in St. Francis and across schools districts throughout the country, the AFT curriculum must be relatively generic and non-specific. Courses such as "Foundations of Effective Teaching" and "Monitoring Student Progress – Classroom Assessments that Work" are required across the board for all new teachers no matter the grade or subject focus. This means that teachers lose the chance to focus on more classroom-specific development. The AFT offers several subject- and/or grade-specific course materials that are also offered in St. Francis, but many of these suffer from a similar issue; since curriculum and assessments are not shared across districts and states, professional development offerings that work across contexts must be divorced from the course content that they are meant to improve upon (Cohen, 2011). Importantly, the AFT's ER&D model offers some means around this difficulty, since the courses are taught by district teachers who are meant to adapt the curriculum to the specific needs of their district's

classrooms. However, teachers' observations in St. Francis about the lack of specificity in Teacher Academy courses demonstrate that this is an imperfect solution at best.

VI. DISCUSSION

Research on the implementation of large-scale instructional reforms provides countless examples of variation in program operations across district or school units due to local adaptation of top-down policy initiatives. This study explores a specific incarnation of this variation as it relates to the scale-up of state teacher improvement systems. In doing so, it offers two insights to the larger body of work on implementation and district capacity.

First, the findings demonstrate the complex interaction between local policy design and the state's capabilities for program oversight and guidance. State education agencies tend to be quite effective at setting up compliance measures for districts to follow, and the Minnesota Department of Education created a series of regulations and guidelines for Q Comp applicants that standardized many aspects of the program. These rules seem to have been relatively binding even though the state's ability to set requirements for districts after they had been allowed into Q Comp was weaker than its authority over initial applications (Schwartz, 2012). At the same time, the state effectively exited from the arena where instructional guidance was concerned, leaving a broad space for districts to define what it meant to engage in instructional improvement as a part of Q Comp. Districts filled in the gap according to their own needs and abilities. Here, the broader story is less about local district personnel co-opting a state policy than it is about districts that were forced to piece together their own understandings about particular areas of a policy in response to high levels of ambiguity. The study demonstrates several of the ways that local policymaking with regard to teacher reform finds its place within the contours of state action.

The ambiguity that districts confronted around instructional reform might be understood partly as a consequence of decisions made at the state level, some which could be altered with the benefit of hindsight, but it is additionally a consequence of the political environment that both the state and its districts inhabit. This environment offers few of the tools such as a unified program of teacher education or aligned curriculum and assessments, all the pieces that Cohen (2011) refers to as the infrastructure, necessary to support strong instructional intervention. Under these circumstances, the theoretical aims of teacher improvement systems – to align teachers’ opportunities to learn with their on-the-job incentives – are likely to fare poorly (Schwartz, 2012). Without a stronger set of resources to call on for the design and development of these programs, it is quite impressive that each of the districts I studied seemed to have made a genuine effort to make teacher learning a meaningful piece of their Q Comp designs.

Second, the study contributes a new concept for understanding variation in district capability around instructional reform. The direction of instructional change depends on the sources of new ideas about practice that teachers have access to, yet policymakers who support teacher improvement systems rarely explicitly discuss where these ideas will come from. This study traces the development and enactment of three different sources of instructional expertise. In one of my case studies, the Q Comp program was designed in such a way as to encourage teachers to learn from other teachers – from their peers – through intensive peer coaching and classroom observations. In another, the Q Comp program was designed to support the analysis of student data with the intent that the new analyses lead teachers take on new understandings about student needs. In the third, the program was designed around teaching content and techniques provided by a contract with an external agency.

In each of these cases, I argued that the ideas about instructional learning embedded within each district's Q Comp design depended in large part on the ways that instructional expertise had been previously defined within the district and the particular demands made on program leaders at the time of the Q Comp application. While this finding might have been suspected from other research on the ways that previous district programs shape engagement with later reforms (Coburn & Talbert, 2006), it is also quite consequential for the development of teacher improvement systems since the different modes of program support different possibilities for teacher development. In Farmington and Centennial, which relied solely on internal resources to drive teacher learning, teachers struggled to determine how to identify and pass on the elements of practice that might help other teachers improve their own classrooms. In St. Francis, teachers were given far more clarity around the elements that they were trying to incorporate into their own practice, which was a point of crucial importance for a program that aimed to raise teacher capabilities with concrete learning opportunities. However, the actual professional development that teachers received tended to be generic and divorced from the details of their own grade levels, subject areas, or teaching context.

Ultimately, this study suggests that without greater guidance around instruction, district incarnations of teacher improvement systems will differ widely in their ability to genuinely support teacher improvement even in the midst of a program that is relatively standardized along a series of other dimensions. One way then of thinking about district capability for instructional reform might be to consider the ways that certain patterns of teacher guidance have already developed in the district before the adoption of a particular intervention, and what the implications of these patterns are for the intervention at hand.

These findings have several policy implications. First, they suggest the need to consider the different types of support for teacher improvement that might develop within the district setting and to determine what possibilities exist for supporting each of these elements separately. It is likely, for instance, that the strongest version of Q Comp would make use of all three models of teacher learning that I described above, looking something like the St. Francis model where external ideas get filtered through the context-specific needs of district teacher-leaders, while correcting for the generic nature of the externally provided professional development. The degree to which this is possible, and the relationship between the state and districts in this endeavor, remains an open question. This article suggests that more state regulations around program structure probably would not have made an appreciable difference, but one could imagine several other roles the state might play encouraging a better balance between sources of internal and external expertise. One possibility would be for the state itself to become the major source of instructional guidance in these programs, yet this would require not only a different sort of focus and organization for the state department but also a different set of expectations about the relationship between state and district governance. Another possibility would be for the state to do more to facilitate the spread of instructional expertise across districts, perhaps working with a set of external agencies to insure that districts can call on these agencies to create the kind of partnerships that I found in St. Francis. A more structured set of partnerships aimed at a particular state program might also help to reduce the overly broad nature of the professional development that I found in St. Francis.

Another implication is that discussions about teacher improvement systems must become more explicitly concerned with the creation of instructional expertise. If, as I have argued here, teachers' opportunities for learning are crucial mediators for the outcomes of such programs and

if these opportunities are highly variable at the district level, we cannot expect to see widespread success unless we do more to address these issues in the program design phase. Thus, states engaged in designing and adopting such programs would do well to consider more specifically the goals of the program with respect to teacher learning and the types of guidance that could support these goals. This essay suggests that all types of teacher learning are not created equally and that different sources of expertise about instruction offer different opportunities for teacher development. Finding ways to support districts in accessing useful sources of expertise should be a central element of teacher improvement system design rather than an afterthought delegated entirely to district adopters.

Finally, the study raises several questions for future research on this topic. One particularly salient question concerns the development of these systems over time. While the districts that I studied had each spent around a half decade in the Q Comp program, my study provided only a single snapshot of the way that people viewed the program. How does experience inform program perceptions and operations? There were some signs, for instance, that both Farmington and Centennial had become increasingly aware in the last couple years of the need to supplement their program designs with outside training, and it is possible that they might move in the direction of St. Francis' model. But it was striking to see how little interaction there was across Q Comp districts around program design. Moreover, due to the increase in the number of Q Comp districts, the state department recently decided to shift its yearly review process from in-person visits to a set of phone calls and document exchanges, further limiting the transmission of ideas across districts. It would be useful to know whether districts in Q Comp (or another similar program) for long periods tend to develop their programs in ways that bring them closer together or farther apart.

REFERENCES

- Baeder, J. (2011). "Rapid change in teacher evaluation." Education Week blog accessed online at http://blogs.edweek.org/edweek/on_performance/2011/04/rapid_change_in_teacher_evaluation.html on July 1, 2012.
- Berman, P., & McLaughlin, M. W. (1977). *Federal programs supporting educational change: Vol. 7. Factors affecting implementation and continuation*. Santa Monica, CA: The Rand Corporation.
- Bonner, S. E., & Sprinkle, G. B. (2002). "The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research." *Accounting, Organizations and Society*. 27: 303-345.
- Burch, Patricia Ellen. (2006). The new educational privatization: Educational contracting and high stakes accountability. *Teachers College Record*, 108(12), 2582-2610.
- Camerer, C.F., and Hogarth, R.M. (1999). "The effects of financial incentives in experiments: A review and capital-labor-production framework." *Journal of Risk and Uncertainty*. 19(1-3): 7-42.
- Chetty, R. Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., and Yagan, D. (2010). "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." National Bureau of Education Research Working Paper No. 16381.
- Cobb, P and Jackson, K. (2011). "Analyzing educational policies: A learning design perspective." *Journal of the Learning Sciences*. 1-35.
- Coburn, Cynthia. (2001). "Collective sensemaking about reading: How teachers mediate reading policy in their professional communities." *Educational Evaluation and Policy Analysis*. 23(2): 145-170.
- Coburn, Cynthia., & Talbert, J. E. (2006). "Conceptions of evidence use in school districts: Mapping the terrain." *American Journal of Education*, 112, 469-495.
- Cohen, D. K., & Barnes, C. A. (1993a). "Pedagogy and policy." In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.
- Cohen, D. K., & Barnes, C. A. (1993b). "Conclusion: A new pedagogy for policy?" In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.
- Cohen, David, K. and Spillane, James, P. (1993). "Policy and practice: The relations between governance and instruction." *Review of Research in Education*. 18: 3-49.

- Cohen, D. K. & Ball, D. L. (1999). *Instruction, capacity, and improvement*. CPRE Research Report No. RR-043. Consortium for Policy Research in Education.
- Cohen, D. K. and Moffitt, S. L. (2009). *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Cambridge, MA: Harvard University Press.
- Cohen, D. K. (2010). "Teacher quality: An American educational dilemma." In Kennedy, Mary, K. (Ed.) *Teacher assessment and the quest for teacher quality: A handbook*. San Francisco, CA: John Wiley & Sons.
- Cohen, D. K., (2011) *Teaching and its predicaments*. Cambridge, MA: Harvard University Press.
- Corbin J., & Strauss, A. (2008). *Basics of Qualitative Research 3e*. Los Angeles: Sage Publications.
- Diamond, John B. (2007). "Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction." *Sociology of Education*. 80: 285-313.
- Eisenhart, M. A., & Howe, K. R. (1992). Validity in educational research. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education*. San Diego, CA: Academic Press.
- Elmore, R. (1996). "Getting to scale with good educational practice." *Harvard Educational Review*. 66(1): 1-26.
- Firestone, William A., and H. Dickinson Corbett. (1988). "Planned educational change." In N. J. Boyan (Ed.), *Handbook of research on educational administration* (pp. 321–340). White Plains, NY: Longman.
- Foley, Eileen, M. (2010). "Approaches of Gates Foundation-funded intermediary organizations to structuring and supporting small high schools in New York City." Policy Study Associates. Accessed online at <http://www.policystudies.com/studies/?id=1> on April 4, 2012.
- Hartley, J. F. (1994). "Case studies in organizational research." In C. Cassell & G. Symon (Eds.), *Qualitative methods in organizational research: A practical guide*. Thousand Oaks, CA: Sage.
- Jennings, Nancy, E. (1996). *Interpreting policy in real classrooms: Case studies of state reform and teacher practice*. Columbia University: Teachers College Press.
- Jerald, C. (2012). "Movin' it and improvin' it: Using both education strategies to increase teaching effectiveness." Center for American Progress. Accessed online at http://www.americanprogress.org/issues/2012/01/movin_it_improvin_it.html on June 29, 2012.

- Kane, T., Rockoff, J., and Staiger, D. (2008). "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education Review*. 27: 615-631.
- Massell, Diane. (2001). "The theory and practice of using data to build capacity: State and local strategies and their effects." In *From the Capitol to the Classroom: Standards-Based Reform in the States*, ed. Susan H. Fuhrman. Chicago: University of Chicago Press.
- McLaughlin, M. W. (1990). "The Rand Change Agent Study revisited: Macro perspectives and micro realities." *Educational Researcher*. 19(9), 11-16.
- Meyer, J., Scott, R., Rowan, B., and Deal, T, Eds. (1992) *Organizational environments: Ritual and Rationality*. Newbury Park, CA, Sage.
- Miles, K. H., Odden, A., Feranich, M., Archibald, S., and Gallagher, A. (2004). "Inside the black box of school district spending on professional development: Lessons from comparing five urban districts." *Journal of Education Finance*. 30(1): 1-26.
- Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. (2nd ed.). Thousand Oaks, CA: Sage.
- National Center on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Accessed online at http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf on June 19, 2012.
- Odden, A. (2008). "New teacher pay structures: The compensation side of the strategic management of human capital." Consortium for Policy Research in Education.
- Patton, M. C. (2002). *Qualitative research and evaluation*. Thousand Oaks, CA: Sage Publications, Inc.
- Pressman, J. L., & Wildavsky, A. (1973). *Implementation*. Berkeley, CA: University of California Press.
- Rowan, B. (2002). "The ecology of school improvement: Notes on the school improvement industry in the United States." *Journal of Educational Change*. 3: 283-314.
- Schwartz, N. (2012). "Aligning Teacher Improvement Strategies at the State Level: The Design and Implementation of Q Comp in Minnesota." Dissertation essay.
- Siemson, E., Roth, A.V., Sridhar, B. (2008). "How motivation, opportunity, and ability drive knowledge-sharing: The constraining factor model." *Journal of Operations Management*. 26: 426-445.

- Spillane, J. P., Reiser, B. J., and Reimer, T. 2002. "Policy implementation and cognition: Reframing and refocusing implementation research." 72(3): 387-431.
- Spillane, J. P., & Thompson, C. L. (1997). "Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform." *Educational Evaluation and Policy Analysis*. 19: 185–203.
- Weatherly, R., & Lipsky, M. (1977). "Street-level bureaucrats and institutional innovation: Implementing special education reform." *Harvard Educational Review*, 47(2), 171-197.

CHAPTER FIVE

CONCLUSION

State education agencies (SEAs) today bear little resemblance to the state agencies of several decades ago. For most of the past century, even into the 1970s, the institutions that on paper claimed constitutional authority over public schooling tended to operate as administrative units that played a small and mostly inconsequential in the greater educational system (Herrington & Fowler, 2002; Murphy, 1973). Standards-based reform beginning in the early 1980s launched a significant transformation, giving SEAs new authority to measure school outcomes and establish accountability requirements. These shifts transformed the ways that many such organizations interacted with local schools and districts, making the SEAs both the ultimate arbiter of educational success as well as the agency in charge of setting consequences for failure and ostensibly resources for success. No Child Left Behind, signed into law by President George W. Bush in 2002, raised the stakes further for state agencies, making test-based accountability the law of the nation and assigning state governments the central role in the design and enforcement of these systems.

Somewhat surprisingly, many of these changes took place amid broader pressures to shrink and streamline the state's administrative footprint. An *Education Week* survey in 1998 found that at least 27 SEAs had reduced the number of employees on payroll since 1980 (EdWeek, 1999). Although I have not found data on SEA hiring post-NCLB, researchers have extensively documented the difficulties that many state agencies faced in implementing the

requirements of NCLB with what these agencies viewed as an insufficient number of staff and insufficient funding (Center on Education Policy, 2002; Center on Education Policy, 2007; Cohen & Moffitt, 2009). Thus, staff and funding shortages reduced the presence of states in districts and schools even as their responsibility for these elements increased.

Perhaps more importantly, the set of standards-based reforms did relatively little to change a broader political environment where the voice of state agencies as a source of instructional guidance often gets drowned out by a host of sometimes contradictory and often incoherent policies set locally by schools, districts, and various private contractors (Cohen & Spillane, 1993). State standards in turn have often developed into political documents that emphasize breadth over depth, offering little to narrow the range of instructional choices (Schmidt, Want, & McKnight, 2005; Finn, Julian, & Petrilli, 2006).

Such pressures based in politics and resources have forced SEAs into a difficult balancing act. On the one hand, SEAs now play larger roles than ever with respect to school policy and practice. On the other, they have mostly relied on localities to make sense of new accountability policies and to determine on their own how to meet the requirements. While there is some evidence that NCLB has had positive impacts on student achievement that appear to outweigh the program's costs, the policy's progression has also prompted considerable backlash from a wide range of constituencies. States have responded by weakening their accountability requirements, by suing the federal government, and by applying in large numbers for waivers around various elements of the law. Reauthorization of the law remains stalled in Congress after five years of political wrangling (Rich, 2012).

While this drama continues to play out, SEAs are now in the process of a second and related transformation, moving into a new realm of educational oversight by way of programs

like Q Comp and the larger RTTT reforms of the past few years. In part, this move represents an acknowledgment that the influence of previous accountability policies has been limited. The new programs, which make state agencies a partner in district human capital strategy, represent an astounding change in expectations for state government and a step that would have been unthinkable several decades ago. Yet these programs still take shape within a context where schools and districts tend to operate quite independently and where local unions have long histories of control over teacher salary and tenure policy. As described above, state agencies have increased their reach in some respects, however they have generally not increased their capabilities with respect to instructional guidance. Thus, this next expansion of state power will again demand new kinds of cooperation between states, districts, and schools and new compromises between programmatic goals, resource constraints, and political pressures. These compromises do not invalidate the enterprise, but they likely place real limits on what might be possible. Understanding these limits and the likelihood of producing robust systems for teacher improvement within these limits is a crucial next step for both researchers and policymakers.

This dissertation illustrates from three different vantage points how implementation of such a teacher improvement system has played out in Minnesota. The first essay (Chapter Two) evaluated Q Comp's effects on teachers' reports of school practice and student achievement, finding that the program had led to only minor and mostly insignificant increases in student test scores, although I do find intriguing evidence that the program has shifted the achievement distribution by increasing the test scores of previously low achievers while lowering the scores of previously high achievers. The second essay (Chapter Three) considered the design and implementation of Q Comp at the state level, describing the challenges that the demands of the program legislation and the state union response placed on the Minnesota Department of

Education and the way that the department coped with these challenges by pairing relatively effective structural oversight with weak guidance and training. The final essay (Chapter Four) focused on district implementation, illustrating how three districts' program design choices around teacher support and development produced programs that differed considerably in their ultimate significance for teachers even as they maintained a rough structural similarity.

Given these findings, what does the work suggest more broadly for state and district improvement policies?

In the previous chapters of this dissertation, I noted several important contrasts between Q Comp and the newer group of Race-to-the-Top-inspired state designs that set limits on the generalizability of this study. Most of the new programs will be implemented across their respective states whereas Q Comp is an opt-in program that has attracted mostly the well-resourced urban and suburban districts in Minnesota. It is difficult to forecast what such programs will look like in new environments that are, on the one hand, likely to be more resistant to the reforms but, on the other, potentially more in need of the kinds of changes that these programs promote. Additionally, the newer state programs tend to go further in the direction of standardized evaluation and pay reform than Minnesota's program. Many such programs will require districts to choose among a smaller set of state-approved rubrics for teacher evaluation, and they will set stricter standards with regard to the percentage of pay increases that are based on classroom-level student achievement (U.S. Department of Education, 2012). Again, it is not clear what these differences will mean to program success. Stronger pay incentives based on student test scores could drive teachers to focus more squarely on the instructional improvement element of the TIS or they might instead overshadow the other parts of the program, creating an

atmosphere of such uncertainty that it distracts from the instructional learning that these programs hope to generate.

Notwithstanding the differences between Q Comp and its successors, Minnesota's example offers useful insight of the development of such policies within the context of the state education system and provides a possible picture of the way that several aspects of these new policies might unfold.

One set of conclusions concerns the choices about program design and implementation made by various groups within Minnesota and the impacts that different choices might have had on the program's development.

At the state level, regulators' attention became focused on the task of interpreting ambiguous legislation and turning it into program requirements that were transparent to district applicants. This was by no means an easy task and when the Office of the Legislative Auditor assessed the program's direction in 2009, its central recommendations were for MDE to further standardize the application and oversight process. Yet necessary though this process might have been, it came at the expense of other possible types of programmatic support.

One thing that was missing from the state's operation of Q Comp was a strong system for building a knowledge base across district subunits about how to make Q Comp most effective. Although there was some effort to facilitate district interaction through a district program peer review process, this process produced little in the way of genuine learning about policy design. Most district personnel had relatively little awareness about the details of other Q Comp programs, and when personnel shared practices they tended to focus on the practices that produced the greatest buy-in rather than those that produced the greatest learning gains. Since the

state had no forms of program measurement and evaluation built into Q Comp, districts focused their efforts on meeting state requirements for funding rather than on program efficacy.

A differently organized state education agency might have worked harder to make Q Comp implementation into an iterative process of program design and evaluation where networks of schools and districts worked together to determine what the state policy had to offer and how it could best support both teacher and student learning. This would likely have required a more purposeful roll-out process than took place in Minnesota where the program was made available to all interested districts less than a month after the legislation was signed. Yet innovative examples of what has been called design-based innovation research are increasingly available, providing useful examples for how to take on such large-scale policymaking in new and untested areas of practice (Penuel et al., 2011; Bryk, Gomez, & Grunow, 2011).

Another challenge for Q Comp implementation was the state's decision to remain almost entirely divorced from district efforts to define what good teaching looked like or to help teachers determine how to get there. To some extent, this is a reasonable choice that emphasized districts' comparative advantage in the realm of instructional guidance as well as the political reality, which included a long legacy of district independence around teaching and curriculum. However the result encouraged variability in implementation, rewarding only those districts that had the initial capacity to make teacher support an important element of the program. Another option might have been for the state to stay outside the realm of professional development while still helping districts connect with other agencies that could provide more of the training districts needed to make good use of Q Comp. The case of the St. Francis school district described in Chapter Four offers a useful lesson here, showing how a district might both retain Q Comp's focus on internal capacity-building while still drawing on external sources for ideas about

teaching improvement. But for St. Francis to make this work required considerable resourcefulness from program planners and demanded a commitment to a model of professional learning that was driven by the district rather than the state program.

The state might instead have done more to support such partnerships with nongovernmental agencies, perhaps even going so far as to create a list of approved agencies similar to the way many states now handle textbook adoptions. This too would have required a substantial shift in the state's way of thinking about Q Comp adoption as well as improvements to the state's analytical capacity, but it might have allowed the department of education to become more of a facilitator, helping districts procure resources to perform the functions that the department was not able or willing to engage in itself. In addition to making such partnerships more available for districts, this state function might have also led nongovernmental agencies to tailor their designs more to the demands of Q Comp and to find innovative ways to deliver the kinds of supplementary assistance that might have made Q Comp a stronger overall program.

To some degree, the state teachers union's progressively diminishing involvement with Q Comp also looks like a missed opportunity. As I documented in Chapter Three, the union devoted considerable resources to district-level training sessions, and seemed at one point poised to take on a central role in helping districts to think about how teacher learning and teacher improvement might look within Q Comp. Instead, a variety of factors including a series of disagreements with MDE, a change in union leadership, and a shifting political atmosphere in Minnesota led the union to disengage from the program and refocus its attentions on the type of political advocacy more traditional for current teachers unions.

One can imagine an alternate scenario where Q Comp propelled the union from its advocacy role into a role devoted to teacher professional growth. We see this possibility at the

local level in some of the places where Q Comp has thrived. In St. Francis, for example, the local president whose union branch now leads the district's Teacher Academy told me that "fifteen years ago, the only time the union ever got involved with anything that dealt with kids would be negotiations and contracts and better language. We were known as that kind of union... The way we look at it now, it's our job to make teachers better. We're not here to fire teachers, that's the administration's job, but we don't want bad teachers in the system." One way of understanding the change described by this quotation was that the local St. Francis teachers union branch transitioned from serving primarily as a worker-protection organization to something that looked more like a teachers' professional association. If a similar transformation had taken place within the state union, the effect could have been to create an institution focused squarely on the problem of teacher development and thus to provide a large-scale supplement to the state education agency as it administered Q Comp.

Each of the possibilities detailed above might have helped make Q Comp into a stronger program for teacher improvement. At the same time, they each also remain limited by several broader elements of the environment in which the program exists. Thus, a second set of conclusions from the dissertation concerns the context of such teacher improvement systems and the limitations that the context place on the development of the new programs.

For example, the state's unwillingness to engage with districts around teacher learning should not be viewed as a choice entirely driven by lack of personnel or a failure of leadership. The task of determining what this engagement might look like and how to do it well was made far more difficult by a diffuse political system in which disparate influences on instruction tend to proliferate—across classrooms, schools, districts, and states—with few avenues for the state to offer genuine guidance. Similarly, while agencies external to the state seem like important

resources for Q Comp districts to turn as they seek to provide their teachers with new opportunities for professional learning, there are real limits as to the kinds of opportunities that these organizations are likely to offer. Because so many instructionally consequential decisions get delegated to the district and school levels, teaching tools provided by private agencies must remain highly generic in order to allow their use within multiple settings. This lack of specificity in turn can diminish their ability to generate the sorts of classroom change that programs like Q Comp are meant to encourage.

To the degree that Q Comp's limitations are seen as a product of the larger system rather than a series of unfortunate choices, the study provides a cautionary tale for other states currently engaged in the move toward programs of this type. Student achievement gains have been small at best and districts' individual experiences with the program differed considerably. At the same time, my quantitative evaluation did not provide a clear indication that certain design models were universally associated with greater program success, leaving a considerable gap in our knowledge about what might make these programs more effective overall.

If Minnesota's experience is taken as a signal of what is to come, it seems difficult to justify the tremendous efforts that many states are now devoting to scaling up similar sorts of systems. This is not to say that other states' experiences will be identical to Minnesota's. Minnesota itself is unique and includes, among other things, a long tradition of local autonomy that added another constraint to state regulation of Q Comp. Other state designs for teacher improvement systems also have features that are not present in Q Comp; for instance, most of the new plans will cover the entire state rather than select districts and they tend to include more standardized guidelines around the measurements for teacher evaluation. Yet these differences alone do not seem likely to address the difficulties around creating robust systems for teacher

learning that I noted in my study of Q Comp. At the same time, the landscape of reform has also changed in important ways since the start of Q Comp, most notably through the adoption of Common Core learning standards across nearly all U.S. states. These standards, which are arguably more demanding and more focused than the standards that previously existed in many states, represent an important step forward in standards-based reform and also an important step forward for teacher improvement systems, since they might clarify the work of both states and districts around teacher support by offering a clearer view of the type of teaching and learning that these agencies would hope to encourage. If well designed and well implemented, the new standards might address some of the challenges that I detailed above, offering a tighter frame for teacher evaluation and making it easier to design tools for instructional learning and feedback that are both rigorous and specific. Whether or not this will happen is an open question. The paradox here is that making the Common Core standards into the kind of tool I am describing would require program designers and implementers to devote exactly the kind of explicit attention to teachers' instructional techniques that have been problematically absent from most previous cycles of state reform, including the teacher improvement systems that I describe in this dissertation.

REFERENCES

- Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). "Getting ideas into action: Building networked improvement communities in education." In M. Hallinan (Ed.), *Frontiers in sociology of education*. Dordrecht, the Netherlands: Verlag.
- Center on Education Policy. (2002). *From the capital to the classroom: Year 2 of the No Child Left Behind Act*. Washington D.C.
- Center on Education Policy. (2007). *Educational architects: Do state education agencies have the tools necessary to implement NCLB?* A report in the series "From the Capital to the Classroom: Year 5 of the No Child Left Behind Act." Washington D.C.
- Cohen, D. K. and Spillane, J. P. (1993). "Policy and practice: The relations between governance and instruction." *Review of Research in Education*. 18: 3-49.
- Cohen, D. K. and Moffitt, S. L. (2009). *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Cambridge, MA: Harvard University Press.
- Education Week (no author listed). (1999). "State agencies take hands-on role in reform." *Education Week*. 18(41): 24-25.
- Herrington, C. D. and Fowler, F. (2002). "Rethinking the role of states and educational governance." In W. L. Boyd (Ed.) *American educational governance on trial: Change and challenges*. Chicago: National Society for the Study of Education.
- Murphy, J. T. (1973). *Grease the squeaky wheel: A report on the implementation of Title V of the Elementary and Secondary Education Act of 1965, Grants to Strengthen State Departments of Education*. Harvard University Center for Educational Policy Research. Report No. HEW-OS-71-132.
- Penuel, W. R., Fishman, B. J., Cheng, B. H., and Sabelli, N. (2011). "Organizing research and development at the intersection of learning, implementation, and design." *Educational Researcher*. 40(7): 331-337.
- Rich, M. (2012). "'No child' law whittled down by White House." *The New York Times*. July 6.
- Schmidt, W. H., Wang, H. C., and McKnight, C. C. (2005). "Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective." *Journal of Curriculum Studies*. 37(5): 525-559.
- U.S. Department of Education. (2012). State Annual Performance Reports. Race to the Top Fund. Accessed online at <http://www2.ed.gov/programs/racetothetop/performance.html> on April 4, 2012.