Striking a balance with school accountability: Design and use of Progress Reports and Quality Reviews in New York City

by

Kelly Anne McMahon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2015

Doctoral Committee:

       Professor David K. Cohen, Chair
       Professor Carla O'Connor
       Professor Donald J. Peurach
       Professor Kathleen M. Sutcliffe

DEDICATION

This dissertation is dedicated to Thomas, Joseph, and Andrew McMahon (Fortier), who remind me to balance my time between things that I think are important and those that really are, and to my parents, who taught me about the distinction.

ACKNOWLEDGEMENTS

Special thanks are due to my committee chair, David Cohen, for his support of this project and my academic career. I am grateful for his willingness to serve as my advisor and for his simple questions that were profound learning opportunities for me and for his kind patience and precise constructions. I am grateful to my committee member, Don Peurach, for his generous willingness to talk and draw about ideas and his work that guides my own thinking, and to Kathleen Sutcliffe for her service, questions and suggestions that ultimately helped me focus my expansive project. I am tremendously grateful to Carla O'Connor for her support of this project and my academic career and for trusting me with EDUC 392.

I wish to sincerely thank all of the members and former members of the New York City Department of Education whom I spoke with - making this project possible. Without their generosity, I would not have been able to conduct this study. In particular, I would like to thank Chris Groll for inviting me to sit at the table and Karen Ames for her support and permission and the rich opportunities they afforded me.

This dissertation owes a great deal to my friends and colleagues who helped me get through graduate school and this project. This includes my writing-group friends, Michaela O'Neill and Lok-Sze Wong and many other cohort members and colleagues who provided guidance, inspiration, and friendship.

I am grateful to my friends in New York City who allowed me to stay with them numerous times over the course of a year. Houseguests in NYC are no small burden. I am tremendously grateful for the hospitality, support and unconditional friendship. Thank you to Noel Wolfe and Katie Stewart, Sarah Litt and Jaimee Nelson, Hedwig Aerts and Jenny Machida, and Rachel Flynn and Mary McCarthy.

I am lucky: I have several close friends whom I could call at 4 a.m. but there are two I will call out here. I am grateful to Juliet Rogers for sharing her wisdom with me (sometimes before I need it and sometimes after) and Kristin Baja for her zealous support

and for being my dearest and most significant cheerleader. Lastly, I want to thank my sister, Kim McMahon, for her courage to make a home and build a family and for sharing so many traditions, memories, and moments with me.

TABLE OF CONTENTS

LIST OF APPENDICES

Appendix

ABSTRACT

In recent decades, people have turned increasingly to accountability to fix schools. Amidst mounting demands for improved student outcomes, states and school districts are trying to determine how best to apply pressure and provide support. Yet, it remains unclear which strategy will produce desirable changes in teaching and stronger schools.

In New York City, in 2002, Mayor Michael Bloomberg and Chancellor Joel Klein set in motion a series of reforms to overhaul New York City's school system. These leaders sought innovation and improved performance – neither of which could be achieved evenly across schools under the former system. They pursued these goals through a variety of instruments including a novel accountability scheme.

In 2007, the New York City Department of Education designed and implemented two unconventional accountability instruments, the Progress Reports and Quality Reviews. In addition, they created a market for the Children First Networks, external support organizations that were meant to provide targeted supports. The networks were held accountable for school outcomes, as well. This unconventional accountability scheme was meant to change outcomes, as well as, the conditions and processes involved in school improvement efforts.

The three studies in this dissertation explore the conceptual underpinnings of accountability as a reform strategy and the case of New York City's novel approach. I conducted an exploratory study of New York City's accountability tools and the networks' use of these tools in the service of school improvement, which draws on data from sixty-eight interviews, more than ninety hours of observation and stacks of internal documents and external reports about the reforms that took place over more than a decade in the country's largest school district.

New York City's accountability scheme involved a novel attempt to balance pressure and support for school improvement from which we can learn about school

accountability as a strategy for school improvement and system reform. The principle lesson from this study is that researchers and designers should avoid looking for whether balance is achieved and consider, instead, how a district attempts to strike the right balance and learn from its efforts.

CHAPTER I

INTRODUCTION

Ideas, as we have seen, are intrinsically standpoints and methods for bringing about a solution of a perplexing situation; forecasts calculated to influence responses.
– John Dewey, 1916

Instead of a "measure, pressure, and punish" model that sets our students, teachers, and schools up for failure, we need a diagnostic, remediate/accelerate model that personalizes instruction, empowers students, involves parents, and provides real feedback to our teachers.
– Georgia's Superintendent of Schools, Richard Woods, 2015

On January 27, 2015, Richard Woods, Georgia's Superintendent of Schools sent a letter to Secretary of Education, Arne Duncan, members of Georgia's congressional delegation and members of the U.S. Senate and House Education Committee to comment on the pending reauthorization of the Elementary and Secondary Education Act (ESEA). In his letter, Superintendent Woods raised concerns about the use of federally mandated tests to "measure, pressure, and punish" schools and educators – a system that Mr. Woods described as a "broken model." In his letter, he made a plea for a new model of accountability. He wrote, "As a nation, we have surrendered time, talent, and resources to an emphasis on autopsy-styled assessments, rather than physical-styled assessments." The analogy suggests that the current model offers information that, like an autopsy, comes too late to save dead patients. By the time the diagnostics show that schools are failing, there are slim chances for recovery.

Instead, as Mr. Woods suggests, we could imagine "physical-styled" assessments that would permit professional educators to diagnose problems and respond with interventions accordingly – offering troubled schools a better chance of survival and restoration of health. Mr. Woods calls for a system that "personalizes instruction, empowers students, involves parents, and provides real feedback to our teachers." In his plea for a different model, Mr. Woods is calling for a better balance between external

conditions like metrics and measurements with schools' internal conditions and how those working in schools deal with instructional improvement.

Mr. Woods is not alone in his critique of conventional school accountability models that were adopted by most states as a response to the federal law, No Child Left Behind (NCLB). Signed into law in 2002, NCLB required states to give annual math and reading tests for students in grades three through eight (and once in high school) to gauge whether students were making progress and to ensure that all students – regardless of race or income were being educated. While there have been modest gains in student achievement since the law was passed, the law remains fairly unpopular.

The wave of complaints about current accountability models suggests that the models are, in fact, a problem that needs to be fixed. Yet, accountability models were introduced as solutions to a set of complex problems that reach into classrooms and state departments (Peurach, 2011). Seen as a way to monitor school performance and move schools towards better outcomes, accountability tries to balance the schools we have with the schools we imagine we should have. Trying to monitor and move quality, the models have, predominately, emphasized a focus on results as a way to change schools from outside in.

Most accountability policies rely on external instruments – things like mandates or rules that aim to regulate internal school quality through an evaluation of students' scores on standardized tests. Instruments "are mechanisms that translate substantive policy goals into concrete actions" (McDonnell and Elmore, 1987, pg. 134). In the case of school accountability policies, the goal has been to improve student outcomes, and in turn, school quality. The policies offer incentives (in the form of rewards and sanctions) to motivate changes in educators' behaviors in schools and issue rules to states and districts requiring them to monitor school performance and support weak schools. Yet, incentives and rules are not exactly concrete, nor are they explicit guides for how to transform weak teaching into stronger teaching. The instruments used by most school accountability policies created a capability gap that many schools were not able to bridge on their own, nor were all states and districts well equipped to provide adequate support to the schools most in need of help (Cohen & Moffit, 2010; McDonnell and Elmore, 1987).

To address the gaps, states, districts, and external organizations have sought to develop guides and supports, including system-level reforms to help schools improve student achievement, but the guides and supports can sometimes be inexact, vague, conflicting, tardy, or insufficient and the systems can appear incoherent, irrelevant, or contrary to principals and teachers, which makes for uneven and inconsistent school improvement. Given the complexity involved in teaching, and the complications such complexity poses for schools trying to improve instruction within – given their unique environments (some helpful, some not), I began to wonder what if accountability instruments were more explicit translations for developing capabilities? Instead of framing poor outcomes as a matter of weak incentives and not enough data, what if accountability models suggested solutions in terms of norms and professional expertise? Is there a way to balance external evaluations with the development of internal capabilities for school improvement?

My aim with this dissertation is to call for a reconsideration of the taken-for-granted view that, if we just hold schools accountable, they will improve. We have continued to turn to accountability to fix schools and solve educational problems, but rarely do we stop to ask whether this assumption really holds, or under what conditions and for whom it holds.

Over the course of several decades, a system of standards, targets, assessments and consequences was created to provide educative guidance to educators from which teachers could align and adapt their teaching based on the standards and what they found when they reviewed students' test scores (Smith & O'Day, 1991). Some suggested this could serve as "systemic instructional guidance" that would allow "…the state to coordinate curriculum frameworks, student assessments, teacher training, and school change around a powerful, coherent vision of curriculum content" (Smith & O'Day, 1991 in Cohen & Spillane, 1992, pg. 131).

But, over the last twenty years, mounting evidence revealed that the instructional frameworks worked for some schools, but not others. A rising tide of testing replaced the frameworks. The result was the emergence of an accountability movement that emphasized results, but was fairly agnostic about the development of teachers' and schools' capabilities.

In response, several external organizations, including some charter school networks, designed strategies to fill the gap (Cohen et al, 2014). While these organizations and some charter networks demonstrated successes with supporting positive school transformation that led to increased student achievement, the legacy of standards-based reform remains mixed (Cohen et al, 2014). "Standards and tests did not translate into common purposes, common professional norms, common curricula, and a common language for diagnosing and solving educational problems, nor did they turn into common practices of teaching, learning and instructional leadership" (Cohen et al, 2014, pg. 176). Rather, the common testing culture with a strict focus on tests brought about by NCLB may have actually distorted school improvement efforts (O'Day, 2002; Hanushek and Raymond, 2004; Jennings, 2012).

Within the last ten years, there have been several experiments with district transformation and school turnaround designed to address the limitations and weaknesses of earlier system-level reforms. Underway across U.S. cities and states, these experiments include efforts to create new accountability models in order to realize better the goal of improving school quality evenly across all schools.

This dissertation includes three distinct essays that explore the idea of alternative accountability models. In the first essay (Chapter Two), I explored the assumptions behind conventional test-based accountability models. I surveyed literature on school accountability and school improvement in order to understand better how research has framed the conceptual connections between school quality, accountability and school improvement. I offer an explanation for why conventional models fall short of their aims and point towards the novel accountability scheme in New York City, which was implemented under Mayor Bloomberg's charge, as worthy of closer consideration.

The second and third essays in this dissertation are based upon a 2013-2014 case study of New York City's school accountability system. This study draws upon sixty-eight semi-structured interviews with current and former NYC DOE central office administrators, superintendents, Cluster and Network team members, and principals; internal documents and reports; and more than ninety hours of observations to understand better the design and use of NYC's school Progress Reports and Quality Reviews. The Progress Reports relied on quantitative measures of student performance and Quality

Reviews were used to evaluate schools' internal conditions for improving teaching and learning; taken together, these tools were used to apply pressure on schools to hit specific targets, while also attempting to enable and orient school and instructional improvement efforts.

More specifically, the second essay (Chapter Three) considers the designs of the Progress Reports and Quality Reviews in New York City. These two accountability tools were part of a series of reforms referred to as the Children First Reforms that Mayor Bloomberg and Chancellor Klein set in motion in 2002. With these initiatives, Mayor Bloomberg and Chancellor Klein sought to change fundamentally how the NYC system served students. These leaders wanted central office to encourage innovation and improved performance – neither of which was possible under the old system, where schools had few reasons to cooperate with central office and could reasonably respond to supervisors by complying with minimum requirements to avoid further interference, or buffering against intrusions all together.

To change how education was delivered to students, Mayor Bloomberg and Chancellor Klein created a strategic plan to change the school system. Towards these ends, Mayor Bloomberg and Chancellor Klein instituted a variety of instruments, including an autonomy-accountability exchange in which all principals were granted autonomy to make decisions about resources and instruction and in exchange, principals were subjected to stricter performance-based accountability that held them, and their schools, accountable for students' progress on standardized tests *and* the organizational conditions for teaching and learning (Childress et al, 2010; Wohlstetter et al, 2012). The overlay of the new system on the old culture, under which principals, teachers, and central office staff could get by with antiquated capabilities and limited knowledge about school-wide instructional improvement efforts and system reform, created all manner of complications.[1]

The complications included the need to develop new capabilities, dispositions and knowledge to perform new practices within different processes and in unfamiliar relationships – all under novel expectations and unprecedented stakes. The success of these efforts depended on learning – individual and organizational learning. The

---

[1] Thank you to David K. Cohen for helping me express this point.

responsibility for orchestrating productive learning fell to designers of the reforms. In a system like NYC, where people possessed diverse sets of capabilities, dispositions, and knowledge bases, the designers of the reforms needed to consider architectures for learning – the scaffolds, artifacts, and interactions that would permit the diverse teachers, principals and central office staff to engage in learning that would permit them to move towards the shared goals – regardless of people's starting points (Stein & Coburn, 2008).

As a way into this story of system reform, I selected to explore two artifacts created by the designers – the Progress Reports and Quality Reviews in order to begin to understand the complications and dilemmas they encountered. I used a framework suggested by Cohen, Peurach, Glazer, Gates and Goldin (2014) that treats designs for improving schools as a puzzle of complex interactions. The puzzle-orientation allows an analyst to consider the complex interactions among four domains: the schools involved, the designs for improvement, the organization responsible for managing the designs, and the environments in which all of these operate (Cohen et al, 2014). I used this organizing scheme to tell the story of Mayor Bloomberg's core strategy, which was based mostly on the use of management principles to spur improvement in instruction. I portray the blueprint for the Progress Reports and Quality Reviews that was created from the complex interactions among the domains.

With this descriptive study, my aims were to capture the dynamic interactions among the domains and reveal how through their efforts to build coherence, develop improvement capabilities at the district and school levels, and target weak schools, Mayor Bloomberg and Chancellor Klein, with the other architects of the reform attempted to contend with complexity in their school system.

The third essay (Chapter Four) draws on a subset of the case study data I collected to investigate the use of the Progress Reports and Quality Reviews by Networks in New York City. The networks were teams of external support providers that partnered with principals to help with the implementation of the Children First Reforms. The networks were intended to replace the old Board of Education, an obsolete bureaucratic structure that was dismantled completely by Mayor Bloomberg and Chancellor Klein. In line with the empowerment structure that preempted the prior structure, principals selected a network partner. By 2013, fifty-five Children First Networks (CFNs) served the city's

schools. Each network provided operational and instructional supports to approximately twenty-five schools. The networks were seen as external to central office and schools, and as such, they operated as brokers who were supposed to provide targeted supports to schools based on each school's needs.

Mayor Bloomberg and Chancellor Klein created a bi-level accountability structure in which schools and networks were held accountable for results. The networks were held accountable under a Network Performance Management system that evaluated networks based on how well their schools performed on the Progress Reports and Quality Reviews. In addition, the networks operated in a public-sector market for their services – meaning that principals could select to affiliate with a different network, if they were unhappy with the service they provided. If networks did not get results, or failed to satisfy their clients, they could be disbanded; the stakes were high for these support providers.

The networks were critical scaffolds for learning; I believed they were key mechanisms in helping schools develop the capabilities needed to operate under the newly designed system that Mayor Bloomberg and Chancellor Klein created. Given this unique school-support arrangement, I sought to understand how the networks used the Progress Reports and Quality Reviews in their efforts to support school improvement. While there was tremendous variation in the supports provided, which was to be expected as this was one of the benefits of this kind of support structure, certain types of supports were more focused on helping schools determine how to use the Quality Reviews as something akin to an instructional improvement infrastructure.

I found evidence that the networks used both tools, the Progress Reports and Quality Reviews to calibrate and organize the support they provided to their school partners. Most networks used the tools to organize themselves and some used the tools to direct their efforts to engage schools in targeted whole school improvement strategies based on each school's needs. This study suggests that accountability can be both a means and ends for school improvement; however, from the varied ways that networks used the different tools, it was also clear that support and improvement can look very different depending on the focus of evaluations. Finally, this study reveals that commitments to instructional improvements are just as, if not more, important as

influences on the development of continuous improvement practices in schools as explicit expectations and support structures.

Taken together, these essays are meant to inform the debates about our "broken model" of accountability and encourage considerations of the relationship between accountability and system reform. Mr. Woods, in his letter to Arne Duncan, suggests that "autopsy-styled assessments" could help, but as several scholars have suggested, and as I found in the review of networks, assessments are only one aspect of the solution. The education system is complex, which means its problems are complex and require complex solutions and require opportunities for people to learn their way into new, or different performances. New assessments given to students may not be enough. Rather, for accountability models to realize the kinds of even and consistent improvement – for all schools and all students, the models may need a complete overhaul in order to address the capability gaps among the adults that have remained stubbornly wide despite significant investments in system level reforms.

References

Childress, S. Higgins, M., Isimaru, A., Takahaski, S. (2010) Managing for results at the New York City Department of Education. In Education Reform in New York City Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Harvard Education Press. Cambridge, MA.

Cohen, D.K., Moffit, S.L. (2009). *The ordeal of equality: Did federal regulation fix the schools?*. Harvard Education Press. Cambridge, MA.

Cohen, D.K., Peurach, D., Glazer, J., Gates, K., Goldin, S. (2014). Improvement by design: The promise of better schools. The University of Chicago Press: Chicago, IL.

Dewey, J. (1916). Democracy and Education: An introduction to the philosophy of education. . The Macmillan Company: New York.

Downey, M. (2015). Georgia school chief to feds: Stop the 'measure, pressure, and punish' approach. The Atlanta Journal-Constitution (AJC). Get-schooled blog. Retrieved: http://getschooled.blog.ajc.com/2015/01/27/

Hanushek, E.A., Raymond, M.E. (2004). Does school accountability lead to improved student performance? National Bureau of Economic Research Working Papers. June 2004. Retrieved: http://www.nber.org/papers/w10591.pdf

Jennings, J. (2012). The Effects of Accountability System Design on Teachers' Use of Test Score Data. *Teachers College Record*. Volume 114, November 2012, 23.

O'Day, J. A. (2002). Complexity, accountability, and school improvement. In Redesigning Accountability Systems for Education. Susan H. Fuhrman and Richard F. Elmore (eds). Teachers College Press. New York, NY.

Smith, M.S., O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (eds)., *The Politics of curriculum and testing* (p. 233-268). Philadelphia: Falmer Press.

Wohlstetter, P., Smith, J., Gallager, A. (2012). New York City's Children First Networks: Turning accountability on its head. *Journal of Educational Administration*. Vol. 51, no. 4, pg 528-549.

CHAPTER II

CONSIDERING ACCOUNTABILITY
AS A MOTOR OF CHANGE IN ALL SCHOOLS

In the minds of most legislators and in the pages of the tabloid press, quality is linked with accountability: if a school comes out well on tests, grades and targets, it must possess quality.

- Maurice Holt, 2000

Do whatever you can, whatever your role, to change the accountability systems in American schooling to make room for the personal as well as the institutional.

- Joseph McDonald, Emily Klein and Meg Riordan, 2009

With the reauthorization of the Elementary and Secondary Act in 2002 commonly referred to as No Child Left Behind (NCLB), the federal government established test-based accountability as *the* way to stimulate widespread improvement in schools. Based largely on a template established by several States in the 1990s, including California, Kentucky and Texas, standards were linked to assessments to push schools to improve. Current debates about the pending reauthorization of NCLB focus on issues of federalism. One side argues to keep the federal mandate for annual testing and the other suggests that states should be able to choose between annual tests and testing once every three years (Rich, 2015). Rarely do the debates consider questions about the relationships between test scores, school quality, and improvement. Do higher test scores really mean higher quality schools?

Evidence suggests that since the passage of NCLB in 2002, student performance especially in math for disadvantaged students has improved and some achievement gaps between white and minority children have narrowed (Ahn & Vigdor, 2013; Dee & Jacob, 2010; CEP, 2007; Hanushek & Raymond, 2004). Yet, more than a decade after NCLB, achievement gaps remain wide and American children perform well below the

competition on international exams. Income inequality in the United States reached a level such that we were labeled the most unequal of all western countries (Saez, 2013; Fitz, 2015). In early 2015, the Department of Education announced that the national high school graduation rate hit eighty-one percent (highest ever); however, reports have shown that nearly thirty percent of students who attend college need remedial coursework (Sparks and Malkus, 2013).   While test scores may be on the rise, it appears the quality of schools is questionable.

To understand better the connections between school quality, accountability, and school improvement, I surveyed literature on school accountability and school improvement. There is not really a line of research that considers school "quality" directly; rather, it is a concept assumed by studies on school effectiveness and improvement. In my investigation, I also drew on my experience as a teacher in New York City during the implementation of Mayor Michael Bloomberg's and Joel Klein's Children First Reforms.   In 2002, Mayor Bloomberg took control of New York City's schools and sought to reinvent the way the system worked.   Part of the Mayor's strategy involved an innovative model of school accountability that relied on new performance evaluations: Progress Reports and Quality Reviews.   These tools were intended to be instrumental in helping transform the country's largest district into, "…not a great school system, but a system comprised of great schools," according to Chancellor Joel Klein.

NYC's strategy was based largely on the premise that accountability should produce improved outcomes for *all* students through better detection and correction of instruction-related problems. Traditionally, instructional problems have been treated as something individual teachers, working behind closed classroom doors, would solve. Instances of failure were seen as products of individual efforts – either of students or teachers.

The NYC model suggested something different: that common failures were not strictly attributable to individuals; rather, school failure was a system failure and demanded a system-level response. The novel strategy developed in NYC was to create an infrastructure of expectations, evaluations and support in order to address weak performance, encourage continuous improvement, and broaden views on school quality. Progress Reports set the top-down expectation that all students were to make progress in

school. Every school in NYC received one of these evaluative reports, which gave each school a grade based on a combination of multiple quantitative measures. Schools in NYC were also evaluated through a Quality Reviews, which were used to set the expectation that schools had to attend to their internal conditions – the systems and structures that schools had in place to strengthen teaching and learning.

With these two tools, the NYC Department of Education was seeking to create a performance management system that would help principals, in particular, and central office recognize and respond to organizational performances of schools. The NYC leaders based this approach on management principles and sought to create a balanced scorecard that would include lagging *and* leading indicators so that information reported about school performance could be useful for strategic planning and management of school and system improvement.

The concept of balanced scorecards is attributed to business management and linked to scholars from the Harvard Business School, Robert Kaplan and David Norton. In the early 1990s, these scholars revolutionized conventional thinking about performance measures when they suggested that metrics could be used to predict performance – rather than used simply to report what had happened. Born from this line of thinking was a management boom that sought to provide mangers with tools that would permit them to control organizational outcomes and link systematically current actions with future goals (Kaplan and Norton, 1992/2007).

When education reformers tried to apply the scorecard concept to education, they could easily point to students' test scores as lagging indicators that characterized schools' past performances. Leading indicators, on the other hand, proved to be more difficult as most data that is collected about educational outcomes relates to student performance (O'Day, 2002). According to business practice, leading indicators should offer information to managers that would permit them to know whether a school was making progress towards helping students acquire the knowledge and skills specified by standards. Taking the analogy further, leading indicators should provide evidence that would raise red flags when schools were not making necessary progress. Yet, research and practice have not agreed on the precise, measurable data that would be attributed to schools. It has proven rather challenging to determine which data would permit

principals or central office to balance oversight and improvement in a complex system like education. Given the technical, political and social nature of educational processes, leading indicators have been seemingly elusive (OERI State Accountability Group, 1988; Supovitz, 2010).

Despite the challenges, some districts like Montgomery (MD), Naperville (IN) and Philadelphia (PA) have tried to experiment with collecting data that could serve as leading indicators - information like student reading proficiency rates and students' enrollment in algebra. Thought to predict whether students were on track towards being college-ready, these aggregates have been attributed to schools. But few districts have used this kind of information in school evaluations attached to consequences (Foley et al, 2008; Supovitz, 2010; Stecher, Hamilton, and Gonzalez, 2003). Rather, the use leading indicators as accountability metrics has remained a theoretical argument (e.g., OERI State Accountability Group, 1988; Oakes, 1989; Selden, 1990; Porter, 1991) – that is until the Children First Reforms in New York City instituted the Progress Reports and Quality Reviews.

The combination of the two tools in NYC was meant to offer a balanced scorecard of lagging (Progress Reports) and leading (Quality Reviews) indicators (Childress et al, 2010). As separate, but complimentary instruments, these tools were meant to enable managers of schools (principals) and the system (central office) to track results while simultaneously monitoring progress in the development of internal capabilities needed for future growth. This was the beginning of an effort to map the relationships between accountability, improvement, and quality.

In this essay, I use evidence from my review of literature, to argue why NYC's accountability tools were unusual and worth attending to more closely. I begin with a discussion of test-based accountability and explain the basic elements of conventional models. In the section that follows, I present evidence to argue why these models disappointingly fall short of our hopes for accountability. More specifically, I present an argument for why, even after years of strict, test-based accountability, which many have seen as an "intervention" that aims to improve all schools for all students –that there is such drastic variability and inconsistency in the quality of educational opportunities available to students. Using organizational change theories, I propose that alternative

models of accountability should consider motors of change for school improvement. I then discuss recent research on routines as in light of these alternative considerations.

In the final section, I turn to New York City's Quality Reviews as a rare and relevant example of a tool used to hold schools accountable for the quality of the routines used to improve teaching and learning. This tool has the potential to fill the gaps left by conventional models. Yet, more research is needed to know whether this experiment, or other models that consider multiple process outcomes achieve better the aim of using accountability as a school improvement intervention (Jennings, 2012; Oakes, 1989). My aim with this essay is to help broaden debates about school accountability and encourage investigations of the links between accountability, quality, and school improvement.

Conventional test-based models

No Child Left Behind was a sweeping federal law that sought to guarantee the quality of education for all students in the United States – leaving no child behind or underserved by its public schools. It mandated that all states had to develop and implement rigorous academic standards in reading and math, which was not a departure from what was happening at the time. By 2001, a year before NCLB was signed into law, forty-eight states had established statewide testing programs, forty-nine states specified academic standards for certain subjects and grades, and twenty-seven states had systems in place to identify low-performing schools (Mintrop, 2004). Similarly, prior to NCLB, most states published school "report cards" that conveyed district and school status based on student achievement results. NCLB codified the need for standards tied to assessments and systems for targeting low-performing schools that would be reported to parents and the public.

The key departures that NCLB introduced included a requirement for states to set annual progress targets to ensure that all groups of students reached proficiency in math and reading by their senior year in high school. School report cards had to report disaggregated data by student sub-groups including income levels, race, ethnicity, special education status, and English proficiency to make sure all students were progressing at an adequate rate. Based on the objective targets that states set, schools had to demonstrate Annual Yearly Progress (AYP) toward statewide targets. The schools that fell behind

and did not meet AYP for consecutive years were subject to various "corrective action" or "school improvement" interventions issued by the state. Schools that received federal Title I funding were subjected to sanctions.

The implementation of NCLB was fraught with problems and marked by the disappointing reality that the goal of 100% proficiency by 2014 was unattainable. In the 2010-2011 school year, only fifty-one percent of schools nationally met AYP (US DOE, 2012). By 2015, forty-three states were granted flexibility waivers from the U.S. Department of Education because states were unable to meet the initial law's targets. Waivers were granted to states that could demonstrate that they had adequate plans for improving educational outcomes for all students. Instead of changing how standards, assessments and reporting were used, the rules for waivers merely added more reporting requirements including one that states needed to track and report graduation rates. States have invested heavily in establishing data management systems to promote more efficient test-based accountability. For example, by 2007, forty-one states had developed data systems with unique student identifiers (Taylor, O'Day, and LeFloch, 2010), which made monitoring and reporting easier and faster.

The logic for using test-based accountability models has remained fairly fixed: that the use of incentives in the form of sanctions and rewards will motivate educators to align instruction and behaviors to meet predetermined standards and outcomes (Felter, 1994; Smith & O'Day, 1990).   These models "take the school as the unit of accountability and seek to improve student learning by improving the functioning of the school organization" (O'Day, 2002, pg. 294). By holding schools accountable, it suggests that the schools would be the primary organization responsible for helping newly motivated teachers develop the necessary capabilities to align their instruction with the standards. Evidence of success would be clear in the assessment data, and when it was not aligning, the suggestion was that schools would be able to adapt and adjust accordingly. States were charged with taking responsibility and providing adequate supports for the schools that repeatedly failed to produce good outcomes. There have been increasing efforts to put pressures on states, districts and schools to make sure all students can read and do basic math, and yet, there is still such drastic variability and inconsistency in the quality of educational opportunities for all students.

In the next section, I offer an explanation about why I think conventional models based on test-based measures made popular by NCLB fall short. We know largely, by negative example, that the models do not initiate continuous improvement practices evenly across schools, and sometimes even within schools there can be wide variations in quality. My explanation focuses largely on the indicators that we use to recognize and measure quality. I focus on these because I believe it is hard to improve quality, if we do not know what we should look for, or how we would know, if we saw it.

What we know about conventional models and moving instruction in all schools

For the past several decades, education scholars have been studying accountability systems and how schools respond to the pressure. Did the schools change the way they were intended to? The quick answer is, yes – some of them did. And some uncommon schools entered the field and were able to make revolutionary progress in urban centers with disadvantaged students. Uncommon Schools, KIPP Academies and Achievement First are some examples of charter networks that developed recipes for success, but these schools account for less than .2 percent of all public schools, and like the New American Schools Development Corporation that preceded these networks (see Bodily, Purnell, Rasmey, and Keith, 1996), there are debates about the scalability of such models. Some of the comprehensive school reforms saw tremendous improvements in teaching and learning for a significant number of schools, but the development of the educational infrastructure the organizations built remains within the networks and schools within those networks.

Overall, there have been small improvements in national test scores (ie., 4th grade math) and others remain the same (i.e., 8th grade reading), but achievement gaps remain wide and actual achievement levels are still quite far from a goal of one-hundred percent proficiency in math and reading. In 2013, only thirty-five percent of eighth-graders were proficient in math on a national assessment and there was a gap of more than thirty percentage points between white eighth graders and their black peers (NAEP, 2013/Center for Education Reform 2014). The United States spends more money per student on education than other countries, but in 2012, according to international

assessments, we ranked 27th in math compared to other industrialized nations. This leads me to ask, why were some schools able to respond to increased pressure, where as so many were not?

There is a growing field of education scholarship that considers similar questions – drawing largely from organizational learning theories to explore the differences in the ways that schools operate as organizations under pressure. Much of this literature seeks to explain how schools, that do not belong to a networked community, attempt to develop the capabilities needed to continuously improve instruction, and the resources or structures provided by districts to promote such development – focusing also on the districts' organizational learning practices (see, Elmore & Burney, 1998; McLaughlin & Talbert, 2003; Tongeri & Anderson, 2003; Hubbard et. al., 2006/San Diego; Honig, 2008; Sykes, O'Day, and Ford, 2009; Datnow and Park, 2009; Stein and Coburn, 2010). The scholars make two important distinctions between organizational learning theories. One category of theories relates to technical aspects of learning, and the other is social aspects. Finnigan and Daly (2012) provide a useful overview of each.

Technical aspects of organizational learning theory

According to Finnigan and Daly, the technical aspects of learning include the process of detecting and correcting problems to improve organizational effectiveness (Argyris and Schon, 1996). When organizations have structures in place that encourage collective capabilities to detect errors, or diagnose problems, the collective can proceed to create solutions that address the problems (Weick, 1999). Learning, and change in the organization, unfolds through deliberate efforts to respond to error with accurate solutions (Fiol & Lyles, 1985).   A second technical aspect is how organizations acquire new information and apply that knowledge. The acquisition of knowledge is often compared to individual cognitive aspects of learning. James March is often credited with the cognitive perspective and his piece with Herbert Simon has served as the focus for several education scholars applying organizational learning to school reform (ie. Honig & Hatch, 2004; Supovitz, 2009). The key processes these scholars focus on include March's and Simon's conceptions of information search, acquisition, integration and assimilation (Childress et al, 2010). These technical aspects of learning "are core to the

work of improvement" as organizations balance exploring for new ideas and refining existing knowledge in relation to novel knowledge (Finnigan and Daly, 2012, pg. 45).

Social aspects of organizational learning theory

It has been suggested that the cognitive perspective frames what people know, or how people know it (Seely Brown, Collins, Duguid, 1989; Cook and Yanow, 1993), but as several scholars suggest, this fails to account for the know-how that resides in the collective and constructed through collective activities (Cook and Yanow, 1993). A second strand of organizational learning is concerned with the social aspect, cultural learning (Cook & Yanow, 1993), or collective mind that occurs as individuals socially construct meaning through deliberate and informal interactions (Weick and Roberts, 1993). In contrast to the cognitive perspective, which focuses on individual learning, the cultural perspective may consider the ways in which individuals interpret and make sense of situations, thereby creating their organization (Daft and Weick, 1984) or the learning that takes place as an individuals learn to become an 'insider' in an organization (Brown and Duguid, 1991).

The social aspect of learning is concerned with underlying relationships that inform how an organization learns and creates co-constructed knowledge (Datnow et al, 2006; Salomon & Perkins, 1998). Cook and Yanow (1993) described the ways in which the making of flutes was a collective activity – as the knowledge to make the exceptional quality flutes did not reside in any one individual, but the organization as a whole. Organizational learning in this view was conceptualized as an activity of the organization – something that had to be practiced by the group. Cook and Yanow (1993) argued that "organizations act." They use these examples to illustrate this point: The Boston Celtics play basketball. The Concertgebouw Orchestra performs a symphony. According to Cook and Yanow, these groups perform in unique ways. No one would suggest that the New York Knicks and Boston Celtics play basketball exactly the same way. This suggests a distinction between knowing and learning. Each team knows how to play basketball, but to play like the Celtics, there is active, collective learning that occurs. The players have to practice and attend to what it means to play as the Celtics (Cook and Yanow, 1993).

According to Cook and Yanow (1993), organizational learning refers to "the capacity of an organization to learn how to do what it does, where what it learns is not possessed by any one individual but by the aggregate itself. That is when a group acquires the know-how associated with its ability to carry out its collective activity, that constitutes organizational learning" (pg. 378). Treating organizational learning as a cultural attribute of the organization as a whole allows scholars to consider the ways in which "learning is embedded in the deeply held beliefs and shared conceptualizations that develop among members of an organization over time as particular understandings and practices evolve through unconscious and regular interactions" (Supovitz, 2010, pg. 709).

While the studies that draw on these this distinction between the tehnical and social – whether they limit to one (Honig, 2004) or combine (Finnigan and Daly, 2012), they offer valuable insights about the conditions, processes and interactions that help, or prevent schools from developing continuous improvement practices. Yet, what I would like to focus on is the distinction itself. The distinction between technical and social aspects of organizational learning explains why some schools are able to respond to pressure and others do not.

Improvement depends on social aspects

"Accountability policies target the technical aspects of learning and improvement, for example, the development of plans and goals for performance, yet, the social aspects are often overlooked" (Finnigan and Daly, 2012, pg. 65). Bryk and Hermanson (1993) made a similar observation when they suggested that in the United States, school policy over looks school's social systems. According to these authors, the social system consists of work flows, interactions, and interdependencies among individuals, the group, and work in particular contexts. Drawing on Bidwell's (1965) image of schools as "small communities," Bryk and Hermanson suggest that the social arena is where we locate the values and tacit understandings that influence how individuals work together on a daily basis, and how they commit collectively to improving those performances. People's motivations and decision-making processes are influenced by the social systems at work in school culture.

But accountability policy aims to influence behavior without acknowledging or attending to the social aspects of schools (Bryk & Hermanson, 1993). Instead, the models reflect an assumption that work-flows, interactions, and behavior can be controlled by regulating behavior with rules and would be observable in student outcomes. This mechanistic view rationalizes work by treating it as discrete tasks accomplished by individuals who are separate from one another. The math teacher teaches math; she only needs to be concerned with her math classes. The English teacher teaches English. And so on. Students are divided by grade level. The rationalization of work is concerned with division of labor into specialized tasks that can be monitored, and in turn, individuals are held accountable for their specialized tasks and results of such tasks. Generally, bureaucratic accountability treats information as a control mechanism and assumes that work improves through better management of information. As a result, "rules and regulations have long served as substitutes for technical knowledge in schools" (Plank and Smith, 2008, pg. 411), or the social arena that influences the acquisition and expansion of technical knowledge.

The relationship between the bureaucratic rules and the internal mechanisms like norms; routines; rituals for detecting and correcting errors; or structures for collecting, constructing and sharing information that actually determine the quality of work is tenuous at best (Popper and Lipschitz, 1998). According to O'Day (2002), "attempts to control individual and group behavior by means of external rules and policies are notorious for their inevitable failure, especially in situations where tasks and environments are complex and ambiguous. Resistance and compliance are the common responses" (pg. 29).

There is an emerging line of research that highlights the role of environments and conditions that influence how schools respond to evidence and improve instruction. Zavadsky (2012), along with other researchers (see Honig et al, 2010), highlight how certain district characteristics influence whether schools successfully adapt and respond to evidence. Zavadsky wrote, "Chronically low performing schools will not improve without the presence of easily accessible data, strong monitoring systems, an overall strategy for benchmarking multiple data points…and a positive and trusting climate that views data as a tool to improve…" (pg. 44). These findings echo other scholars' findings

that district conditions matter tremendously for schools trying to improve. For example, McLaughlin & Talbert (2003) argued that a district's approach to system reform, degree of coherence and focus on teaching and learning, stance on professional development, and data-based inquiry and accountability effect schools in those districts and their efforts to improve instruction.

A district's culture can affect the kinds of responses that schools select when they respond to evidence, or data. For example, Firestone and Gonzalez (2007) highlight that some districts have an "accountability oriented culture," where as others have an "organizational learning culture." In both cultures, schools are called on to respond strategically to evidence of problems, but accountability-oriented cultures tend to produce short-term, compliance based responses, and organizational learning cultures rely on professional communities and invest in long-term change processes. According to these authors, a district culture that stresses more organizational learning, "…is more conducive to educational improvement than one that stresses accountability, although how these two aims are combined matters" (pg. 153).

Evidence suggests that teachers and administrators do respond to external accountability pressures, but the results are not always positive or clear. For example, classroom-level research points to changes in scheduling (Marsh & Ikemoto, 2006), increased resources for tested subjects (Diamond & Spillane, 2004), and increased focus on test-based instruction (Kane & Staiger, 2002; Murnane, 1987). Principals in New York City in Shipps (2010) study reported feeling overwhelmed by accountability pressures and "many described the need for district support services and external resources, which are not yet available" (pg. 32).

Research suggests that there is reason to be skeptical of the solutions that schools may come up with – especially when a school with low capabilities faces high-stakes for drastic improvements in performance. Scholars warn that there may be unintended consequences that arise, as schools feel compelled to game or cheat the system (Koretz, 2009; Kane & Staiger, 2002; Linn, 2000; Murnane, 1987). Jennings (2012) refers to this kind of solution as distortive data use. A recent example of distortive data use is the 2010 Atlanta cheating scandal in 2011 (Rich, 2013). In Atlanta, Superintendent Beverly Hall, a former Broad Prize winner, and thirty-five staff including teachers and administrators

were charged with thirty-five counts of fraud for altering students' tests to increase scores. A common theme emerges from the scandals and studies showing that policies interact in complex ways not often "predicted by policy makers theoretical (or ideological assumptions)" (Shipps, 2010, pg. 3).

Conventional test-based schemes operate as if external policy will influence the internal development of organizational learning capabilities through the application of pressure for change. While pressure is necessary, it does not provide sufficient influence over the social aspects of schools where valuable mechanisms for organizational learning reside (Massell and Goertz, 2005; Finnigan and Daly, 2012; O'Day, 2002). Finnigan and Daly (2012) found that the "structures and quality of social relationships within schools and district-wide play a crucial role in schools' capacities for organizational learning and improvement" (pg. 65). These findings extend others that schools with collaborative or trusting cultures are more likely to show signs of improvement, and possess abilities to respond to external demands for performance (see Bryk & Schneider, 2002; Mintrop & Trujillo, 2005).

Richard Elmore and colleagues suggest that the mechanisms that influence the social aspects inside schools are found in schools' internal accountabilities. According to the authors, internal accountability mechanisms include individual conceptions of accountability: To whom am I accountable? And shared expectations: For what are we accountable? Lastly, the looked at consequences: How are we accountable? Together, these normative elements comprised a school's internal accountability system.

In a large-scale study through the Consortium for Policy Research in Education, Elmore and colleges investigated how schools responded to external accountability pressures. The authors found that the strength of schools' normative cultures, or their internal accountability systems varied greatly and influenced whether schools were positioned to respond to external pressure. The default in most schools relied on personal discretion to be the main driver of internal accountability. In these cases, there were few shared expectations, or formalized accountability mechanisms to hold teachers accountable to some normative expectation about how teachers should do their work; in this culture, teachers made decisions independent of one another and there were few structures to promote collaboration. This echoes other descriptions of schools as cellular

places, where teachers' work and problem-solving is autonomous – a closed-door version of teaching (Lortie, 1975). Without a sense of shared expectations, or internal mechanisms for holding people accountable, social interactions are not a source of professional learning or school improvement. According to Elmore and his colleagues, in a school where individual responsibility reigned supreme, there was little alignment between a school's internal system and the external environment; teachers did not make sense of the external environment as it related to their school, and in some cases, the external rules influenced little of what they believed they were responsible for. In these types of schools, there were few bridges available to create alignment between the internal and external accountability systems.

The schools with the most complex formulation of accountability were more rare in Elmore's study. The authors reported that complex schools were places where the, "collective expectations gelled into highly interactive, relatively coherent, informal and formal systems by which teachers and administrators held each other accountable for their actions vis-à-vis students" (Elmore et al, 2004, pg. 193). While teachers and administrators in these schools were able to describe and interpret the external demands, the external accountability did not influence their construction of their internal accountability systems.

Based on their study, the authors concluded that, "in most cases, teachers and principals viewed external accountability like the weather – something that might affect their lives in some way, something they could protect themselves against, but not something they could or should do much about" (Elmore et al, 2004, pg. 197). They extended this by writing, "Conditions within schools are logically and empirically prior to the conditions outside schools when constructing a working theory of educational accountability" (pg. 198).

What Richard Elmore and his colleagues found was similar to what Newmann, King and Rigdon (1997) concluded when they compared school's internal accountability as it related to strong external accountability. These authors also saw internal accountability as normative practices in schools. Normative conditions included teacher skill, distributive decision-making, collaboration and shared commitments for clear purposes related to student learning. They found that "external accountability alone

offers no assurance that a school faculty will have adequate technical knowledge and skill, sufficient authority to deploy resources wisely, or shared commitment to a clear purpose for student learning" (Newmann, King, & Rigdon, 1997, pg. 62). Strong internal accountability tended to be accompanied by a stronger organizational capacity to respond to strong external demands for performance, but strong external demands did not spur schools to develop stronger internal conditions.

Several types of reformers emerged to address the deficiencies of standards and test-based reforms that disappointedly failed to trigger necessary changes in internal conditions or practice that would have aligned efforts to improve instruction with the broader goals of test-based models. These reformers sought to target the seemingly weak internal capabilities in some schools – specifically, the lowest-performing, and high-poverty schools that were overwhelmed by the tremendous challenges that exist in trying to manage the complexities of instruction and its improvement in unhelpful conditions.

One line of research comes the Study of Instructional Improvement that examined large-scale comprehensive school reforms that sought to help schools develop collective capabilities to improve instruction by creating strategies that targeted school culture, routines and curriculum, and professional development programs (see Rowan, B., Camburn, E., Correnti, R., and Miller, R. 2009; Peurach, 2011; Cohen et al, 2014). The reforms spanned decades, as did the research, but the key findings related to the efforts of external organizations to build educational systems that directly shaped schools' internal coherence and instructional systems (Cohen et al, 2014). These organizations sought to improve high-poverty schools specifically – an ambitious goal rife with challenges.

Cohen at al, 2014 summarized that the organizations, namely, Success for All, America's Choice, and Accelerated Schools Project, sought to create "coordinated, coherent relationships that enabled communication among schools, and between schools, and the national centers and regional training organizations that they established. These national centers…were vehicles for building common culture as well as professional knowledge and skill…They were professional education agencies that helped teachers and school leaders learn how to turn the designs into practice" (pg. 172). According to the authors, the construction of these organizations, as many were building their own capabilities and designs while trying to engage schools in developing more coherent,

coordinated structures for instructional improvement, was a timely, expensive and expansive endeavor.

In the U.S., there are few models for designing these kinds of systems, and even fewer resources – especially for the development capabilities in environments like urban, high-poverty schools that face tremendous social and economic challenges. Cohen et al (2014) suggest that these organizations developed "educational capability" by building an infrastructure and networked communities through which schools made progress with school improvement – unmatched by local districts and State systems that lacked the kind of infrastructure that the organizations built over time (Cohen et al, 2014).

Similarly scholars have investigated successful charter networks that were able to develop shared systems of professional values, norms, and expertise within individual schools and across a network of schools. Rosenberg (2012), drawing on the work of several scholars (i.e., Byrk, 2009; Raudenbush, 2009) proposed a set of considerations for understanding how charter networks develop instructional systems – not scripted performances, but a knowledge base that guides professional judgment, practical adaption, social routines – norms for practice and improvement (Bryk, 2009). In her examination of Achievement First, Rosenberg found that the charter network "developed an impressive range of instruments and arrangements to support teaching quality across the network that reflected both individual and systemic strategies for teacher and teaching quality support and management" (pg. 180).

These two lines of research – on charter networks and comprehensive school reforms suggest that the ability of schools to develop coordinated and collective responses are directly related to their membership in an educational system that informed professional development, the kinds of instruments teachers had available to them to solve problems of practice, as well as, the quality and effectiveness of teachers' professional expertise and learning communities (Cohen, 2011). Yet, not all traditional public schools belong to one of these systems. Rather, in 2007, there were more than thirteen thousand schools (nearly 15% of all schools – Title I and non-Title I) that were identified as in-need of improvement (Taylor, O'Day, Naftel, and LeFloch, 2010). There are still plenty of schools that need more explicit guidance about developing collective capabilities and addressing the social aspects of organizational learning.

Test-based accountability, as it offers assessments connected to standards for what students should know and be able to do should have helped schools develop coherence in the fragmented U.S. system (Smith & O'Day, 1993), but it may not have lived up to its potential. There have been many positives that have come from greater prevalence of data about achievement gaps among students including, some might argue, the rise of charter schools and comprehensive school reforms that developed instructional systems as described above. Similarly, the development of the Common Core Learning Standards that arose in response to some of the gaps exposed by NCLB shows great promise for the development of stronger instructional systems. However, if we were to treat current test-based accountability as a system-level reform, we could reasonably conclude that it fell short of meeting its aims to trigger consistent and even improvements in teaching or learning for all students.

Based on these older studies of schools' internal responses to external accountability, and more recent studies that suggest district culture, or membership in some networked community that provides schools with an instructional infrastructure affects the degree of variation in schools' organizational capabilities to cultivate strong normative cultures, shared professional expertise and commitments to improving instruction, we can see that external accountability models that rely on models based on test scores are fairly limited in their ability to change norms and professional practice in schools. Rather, it appears that schools with strong internal accountability systems, or those that operate in a district with an organizational learning culture, or belong to a networked system, are better positioned to respond to external pressure for performance. Given that networked communities are unlikely to take over all traditional public school districts (although New Orleans stands as an exception, and current portfolio management models suggest it could happen), what could reasonably help schools develop internal, collective norms and processes for instructional improvement?

In the next section, I use organizational change theories to propose that accountability models could be useful resources for developing internal, collective norms, and processes for instructional improvement. However, theory suggests that the designs for alternative models should attend to *motors of change*.

Organizational change theories

Organizational learning theories have been very useful to education scholars; these theories provided instrumental conceptions to consider how schools respond (adapt or learn), and the challenges that accompany such efforts. The studies that draw on these theories offer tremendous insight into the persistent variability that occurs as schools attempt to respond to accountability policies – explaining pockets of success and why some schools cannot move off the sanctions list. Yet, there are few studies that offer suggestions for how to modify accountability models so that they might successfully spur deliberate efforts to strengthen the normative aspects of *all* schools and begin to influence the quality of instruction evenly across schools. I think organizational change theories hold some promise towards this end.

In the text, *Institutional Change and Healthcare Organizations*, organizational studies scholar, W. Richard Scott with several colleagues tells the story of the significant transformations that occurred in health care systems over the past fifty years. They studied the change processes to explain the changes in types, numbers, and activities of healthcare organizations. They sought to understand why some health care organizations thrive during turbulent policy shifts, where other forms suffer tragic ends. Seeing some important parallels between health care and education, I turned to these authors' consideration of theories on organizational change.

The authors draw on a seminal typology constructed by Van de Ven and Poole (1995) that organizes the four central arguments about organizational change that have appeared over the years. The four include: (1) life cycle theories, (2) teleological theories, (3) dialectical theories, and (4) evolutionary theories. Each type is based on what the theories suggest is the source of change. For example, life cycle theories rely on a metaphor of organic growth and seek to classify stages in an organization's development from birth to death. By comparison, teleological theories assume that an organization is "purposeful and adaptive, directed toward some goal or desired end state. Development is viewed as an iterative process and goal setting, implementation, evaluation, and goal modification" (pg. 62). Dialectical theories treat change as stemming from conflict in a pluralistic world filled with competing forces. Evolutionary theories assume that change proceeds through cycles of variation in which "new elements

(rules, roles, organizations) arise through random change; selection occurs primarily through the competition for scarce resources" (Scott et al, 2000, pg. 63). These types are not mutually exclusive as actors (individuals, organizations, and fields) may go through various types depending on external factors, or the goals each pursue may conflict, or change over time. The value of the typology is that it offers a way to identify different assumptions about the source and nature of change. The authors outline several distinctions; I am going to focus on two.

Internal versus external sources of change
Scott and colleagues suggested that recognizing the "motors" of organizational change is important to differentiate among theories, and in doing so, change becomes predictable (or in the very least observable to the researcher). Evolutionary theories stress external causes of change where new elements may appear randomly requiring an organization to attend to new dynamics. Examples include institutional theory, or population ecology. By contrast, teleological theories emphasize factors internal to the organization and tend to assume that actors are rational decision makers. Examples include strategic management approaches that stress internal actors and processes (Scott et al, 2000, pg. 63).

What is helpful about the distinction between external and internal sources of change, which may at first seem obvious is that "organizations and participants are not passive pawns of external events, allowing external forces to freely reshape them, but take steps to control, modify, and challenge these forces" (Scott et al, 2000, pg. 64). Education policy scholars consider this dynamic, but often refer to the directions as top-down, bottom-up, or mutual adaptation, which Datnow and Park (2012) named the bi-directionality of influence on policy and its implementation. What this line of thinking brings to mind is the extent to which policy, as an external force, and the instruments it bears, can directly influence the behavior of actors who work together in an organization and the culture that they create as they do that day after day.

The distinction between external and internal sources of change is an important one for considering the origin of the change process. External accountability presumes to be a source of change by setting standards and providing feedback on performance, as

well as, offering a set of incentives to change teaching practice. This parallels what Beckhard and Harris (1987) refer to as a "demand system" – "meaning forces in the environment that initiate the process of organizational change" in transition theory (Burke, 2014, pg. 178).

Transition, described by Burke, involves three distinct phases, "the future state, where the leadership wants the organization to get to, the present state, where the organization currently is, and the transition state, the set of conditions and activities that the organization must go through to move from the present to the future state" (pg. 178). According to the theory, to move through, what may not be a "neat or sequential" process of transition, when organizations get to the transition state, they require "transition management…a process of conducting activities such as planning a road map for the change effort" (Burke, 2014, pg. 179). Applying this concept to accountability, we can begin to see that this suggests as schools that face a "demand system" and forces in the environment attempt to initiate change, there should be a road map that makes explicit the activities and technologies needed to transition to the future state.

Current test-based accountability models are fairly agnostic about transition activities and leave it to the professionals to determine. In some instances, this becomes the space where networked communities, or charter schools, or districts step in to offer guidance through an instructional system. Yet, in the absence of an instructional system, or networked community upon which a school might draw, there may be few resources within a school that would enable or support the development of the road map or the transition from current to future state. This suggests that a demand system that made explicit particular, or developmental stages of the transition process could prove helpful for some schools – especially those that lack access to the guidance located in an instructional system.

Incremental versus discontinuous change

Scott and his colleagues highlight the importance of considering the magnitude or scale of changes observed. Like learning, change can occur in episodic rather than a gradual, or linear progression. In education, policy scholars have also described implementation as being a less linear trajectory; rather, it can be more dynamic - non-linear and social endeavor (Honig, 2008; Datnow and Park, 2006).

Scott and colleagues draw on population biologists' theories to highlight that change can occur in "punctuated equilibrium" (Eldrege and Gould, 1972 in Scott et al, 2000). Periods of change are 'punctuated' by periods of turbulence. Change that occurs during major disruptions in the normal order of things is called discontinuous change, or metamorphosis; where as smaller changes that occur during periods of equilibrium are incremental, or adaptive changes at the organizational level (Scott et al, 2000).

Schools vary in their internal, normative environments (Elmore et al, 2004). The kinds of structures, cultures, and systems that schools use to respond to external pressures, or interventions – the kind change demanded by new rules or new resources, including possible partners could play out as incremental changes for some schools given the qualities found in their normative environments, or the same intervention could produce tremendous turbulence in a school's normative environment and require a school to undergo a complete metamorphosis. Current test-based accountability policies do not account for distinctions in the magnitude of change that some schools must go through in order to meet the expectations of raising student achievement, nor do they provide supports for the varying experiences.

Accountability policies aim to change behavior. Although, "as any observer of school reforms will tell you, it is easy to change the buzz words but difficult to change behavior" (Hannaway and Stanislawski, 2005, pg. 58). Rather, theorists have argued that there are three mechanisms that shape behavior: incentives, routines or standard operating procedures, and norms or culture within an organization (Hannaway and Stanislawski, 2005). Conventional accountability models focus on the first – incentives, and leave the other two mechanisms unspecified and open for interpretation. The incentives are useful for applying pressure and creating a demand system, but less helpful as they fail to provide requisite road maps that outline the sets of conditions and actions that some schools need specified in order to be able to make progress towards a more desirable state.

With mounting evidence about the strength of instructional systems and other structures that influence the cultural aspects and standard operating procedures in schools, we can begin to imagine different accountability models that specify target outcomes and conditions, or procedures. Similarly, the models could better reflect the

magnitude of change expected with developmental indicators that reflect the process of change, rather than operating from the base assumption that it will happen and be discernible in student outcomes. If education had a stronger professional base, it could provide better guidance for developing stronger normative environmental criteria, but in the absence of that, the accountability mechanisms could contain normative criteria coupled with rules and regulations that make expectations for adults' performances more explicit. In the section that follows, I provide an example from research of an enabling mechanism that holds promise.

Internal motors of change: Routines

 Education and institutional scholars have used a metaphor of schools as loosely-coupled systems to explain why policy often fails to reach inside classrooms (Fusarelli, 2002). Coupling refers to the relations and interdependency between levels within a system or organization. How loose, or tight the coupling is suggests how changes in one level affect another.   The loose-coupling metaphor was often applied to schools to suggest how administrators often managed the organizational conditions around instruction, but rarely dealt directly with it. In many instances, the administrators were seen as buffering instruction from external pressure – leaving policy makers frustrated because they had few channels through which they could influence the technical core of schools (Elmore, 2000).

Drawing the coupling metaphor as something organizations do, rather than just have, Spillane and his colleagues (2011) sought to understand how school leaders responded to a changing institutional environment – with stronger accountability and greater specification of standards (i.e., Common Core Learning Standards) that call for more ambitious forms of teaching.   The authors found that school leaders created organizational routines in their schools that were used to "standardize curricula, monitor student and teacher performance, and make classroom practice more transparent" (Spillane et al, 2011, pg. 600).   More specifically, these leaders set clear expectations for staff, standardized the instructional program, and created processes for monitoring results that they used with staff. More regular monitoring was used to identify problems in the technical core and apply solutions, as needed.

Leaders in these schools used student achievement data to motivate teachers to see the value in making classroom instruction more responsive and less distant from state and district regulations. The researchers concluded that these routines, while they were described as creating, "a dramatic shift in ways of doing business at their schools" (pg. 600)," and as such offered tremendous promise, the findings are also quite limited.

The findings are limited because this study was conducted in four elementary schools. The way elementary schools are organized make it easier to target "grade level" and "content area coaches" as useful locations through which a leader can seek to move towards consistent and standard curricula and assessments within and across grades. It is easier to make teaching transparent when there are fewer teachers and its easier to set expectations for adult performances. It is less likely that principals of high schools would be able to use the strategies that principals in this study used (even less likely in large high schools where elective courses and teacher specialization pose significant hurdles to standardizing an instructional program).

Secondly, leaders in these schools encountered significant pushback from veteran staff in their schools. The leaders had to be able to overcome and persist through conflict. One of the principals in the study suffered such pushback that the district conducted an investigation into the principal and her efforts to change the formal structures in the school and make teaching transparent. The principal survived the investigation, "though conflict persisted" (Spillane et al, 2011, pg. 604). To pursue a strategy to use routines to change culture and practice, school leaders would need to be fairly sturdy – that is, in the absence of a networked community that could offer resources to manage the conflict that arises during cultural transformation.

Despite these challenges and limitations, Spillane and his colleagues suggest that this study suggests that "organizational routines offer a particular way of thinking about school reform in that the development of practice (i.e., administrative practice) is the central focus, as distinct from a central focus on developing the knowledge of one or more school leader…practice is defined as interactions among school staff (pg. 615). Routines, according to the authors, are a way of influencing these interactions. The schools in this study developed organizational capabilities and changed their normative environments by using new routines to promote collective and collaborative responses to

their external environments that demanded improved student outcomes. Routines became a way in which school leaders were able to embed core ideas advanced by government regulation into the formal structure of the school, and in turn, part of a school's social culture as staff start to take them for granted as part of their regular work (Spillane et al, 2011).

Panero and Talbert (2013) similarly call for a focus on routines as a motor of change and continuous improvement in schools. These authors looked at a strategic inquiry model, the Scaffolded Apprenticeship Model (SAM) program, which involves teams of teachers using a structured protocol and process that moves through various phases of inquiry-based strategies for, first, improving outcomes for specific students; second, teams' decision-making based on evidence; and three, extending shared accountability and collective problem-solving across a school (Panero and Talbert, 2013). The authors found that in schools that used strategic inquiry routines, and the use of the strategic inquiry, "shifts schools toward evidence-based culture"; in these schools, teachers' typical beliefs and norms changed to support the development of continuous improvement capabilities (pg. 5). For example, "team after team has discovered that they cannot make assumptions about what struggling students know and do not know. They have discovered that students not knowing these skills is a result of their not being taught (as opposed to students not being able to learn them), and that when teachers teach these skills, students become engaged, learn, and make giant leaps in their progress" (Panero and Talbert, 2013, pg. 5).

Taken together, research from Spillane et al (2011) and Panero and Talbert (2013) suggest that routines pose a powerful mechanism through which schools were able to transform cultures from ones that locate professional knowledge within individuals to cultures that encourage the development of professional knowledge and expertise through regular, structured, social interactions. These studies suggest that routines were one way to develop responsive, internal capabilities and norms focused on shared conceptions of instructional improvement. These suggest that it is worth attending to the use of routines as a positive motor of change inside schools that did not belong to networked communities, or have access to an instructional system.

This raises questions about what kinds of routines would be helpful? What could an external policy name as valuable indicators that would help leaders make legitimate changes that move beyond symbolic routines to embed new formal structures that encouraged authentic interactions and collective problem-solving around instruction related problems? The Quality Reviews in New York City offered some possible answers.

Learning from New York City's Quality Reviews

In their book, *Strategic Inquiry: Starting Small for Big Results in Education*, Panero and Talbert (2013) describe strategic inquiry and how it works. They explain that the model of strategic inquiry that they studied has been adapted to district and school turnaround efforts in Oakland, Boston, and Rochester; however, they explained that they selected to focus specifically on New York City's version, where Inquiry Teams had been an important feature of the Children First Reforms in 2006. "The track record of New York City schools that implemented the model for several years shows significant growth in teachers' leadership and evidence-based decision making, along with steady gains in student achievement" (Panero & Talbert, 2013, pg. 10).

The Inquiry Teams in New York City were one part of the Quality Review rubric, which was used to evaluate all schools on their normative environments, conditions, and processes for strengthening teaching and learning. By 2010, the Quality Review Rubric specified five indicators (each with sub-indicators) for evaluating schools' organizational conditions, including routines. For example, indicator 4.2 states that schools should "Engage in structured professional collaborations on teams using an inquiry approach that promotes shared leadership and focuses on improved student learning."   The DOE defines an "inquiry approach" as one where "teacher teams systematically analyze key elements of teacher work including classroom practice, assessment data, and student work for students they share or on whom they are focused, resulting in shared improvements in teacher practice and mastery of goals for groups of students" (NYC DOE QR Rubric, 2015). Because the Quality Review specifies and makes explicit which routines should be evaluated and is used to hold all schools in the city accountable (attached to very high stakes), it offers a unique opportunity to learn about an alternative accountability model that aims to combine external pressure for improving student and

adult learning with enabling supports that offer guidance for how to change internal conditions.

In 2007, New York City implemented the Quality Review to serve as a qualitative assessment to match its newly implemented Progress Reports.   Together, these two tools were meant "to strike the balance between instilling accountability for past results and encouraging continuous improvement toward future outcomes" (Childress et al, 2010, pg. 88).   Seen as offering leading indicators of performance, the Quality Review process is a two-day school visit conducted by an expert educator who observes classrooms, looks at student work and meets with teachers, families, students, and the principal to better understand the qualitative aspects of the school. Using a rubric that looks at three domains: instructional core across classrooms, school culture, and structures for improvement, the evaluator, using her observations, as well as, information like the school's self-evaluation, provides detailed feedback and rates the school along the various indicators in the domains. The purpose of this process is to encourage school communities to develop shared understandings about collaborative inquiry, improvement planning, and school support (DOE, 2013).

As of 2007, all schools in New York City receive a Quality Review at least once every four years, and more if the school has been targeted as a priority or focus school by the New York State Board of Education. Between 2007 and 2013, schools' ratings on the Quality Review were used, in combination with Progress Report grades, and other information, to determine school closings, principal tenure and merit pay.

The Quality Review in New York City offered a new way for educators, policy makers, and the public to recognize and monitor school quality. In addition, it provided a potential motor of change that looked specifically at a school's routines and organizational conditions as they related to instructional improvement – creating a potential instructional system for all schools. To date, we lack knowledge of the usefulness of this tool, which was described by Childress et al (2010) as an "organizational learning tool." More research is needed to understand the potential of this tool to fill the significant gaps that currently exist between what we hope school accountability will accomplish and what it does.

Conclusion

In this chapter, I began by suggesting that perhaps there is a reason we may want to question conventional thinking – the taken-for-granted ideas about the linkages between school quality, improvement and accountability models that rely predominately on student test scores. I proposed that the reforms in New York City under Mayor Bloomberg – between 2002 and 2014 offered an unusual case where there were a set of disruptive ideas that stood to challenge conventional thinking about the cause and effect relations among school accountability, school improvement and quality.

In the sections that followed, I provided a brief description of conventional models to provide a basic overview of my understanding of conventional models. Drawing on research, I offered an explanation for why conventional models fall short of evenly achieving higher levels of school quality and improvement across schools. In an effort to critique and construct, I argued that attempting to measure and evaluate the routines performed by schools could be a useful target for managing change and improvement efforts in an accountability scheme.

Based on my experience with and knowledge of the unique accountability tools in New York City, I suggested that Quality Reviews provided a rare opportunity to develop foundational knowledge about an external accountability scheme that attempts to monitor, measure and influence schools' routines for continuous improvement. More specifically, New York City's Quality Review presents an unusual opportunity to examine a district's attempt to influence the development of continuous improvement capabilities through explicit evaluations of how schools manage the social aspects and conditions for instructional improvement.

The reforms in NYC are, most notably, ambitious attempts to measure schools' internal conditions as they relate to strengthening teaching and learning, but they are also worth attending to because of their scale. Unlike other experiential reforms happening in other urban centers like Chicago, Boston, Philadelphia, where schools were being granted autonomy and provided different supports to develop internal capabilities, the reforms in New York City, under Mayor Bloomberg, were system-wide and targeted *all* schools. Reform at this scale is rare as twenty-eight of the 50 states in the U.S. manage fewer schools than the New York City Department of Education (Parthenon, 2013). The case

of accountability reforms in NYC offers a rich opportunity to understand how to bring about change consistently and at scale.

Given the current climate in which there are significant efforts to develop national standards for learning matched by the rollout of new assessments tied to those standards, yet no clear accountability model that helps teachers develop capabilities to meet higher standards for teaching and student learning, it is an ideal time to consider alternative models and measures that make explicit what adults in schools should know and be able do in order to achieve more collaborative, coherent, and collective instructional improvement practices capable of influencing student learning in positive ways.

References

Ahn, T., Vigdor, J. (2013). Were all those standardized tests for nothing? The lessons of No Child Left Behind. American Enterprise Institute. May 2013.

Anderson, S.E. (2003). The School District Role in Educational Change: A Review of the Literature. International Centre for Educational Change (ICEC) Working Paper #2, August 2003.

Anderson, S.E., Leithwood, K., Strauss, T. (2010). Leading Data Use in Schools: Organizational Conditions and Practices at the School and Districts Levels. Leadership and Policy in Schools. 9:3, pg. 292-237.

Bayne-Jardine, C., Hoy, C., Wood, M. (2000). Improving quality in education. Falmer Press: London.

Bryk, A. S., Hermanson, K.L. (1993). Educational indicator systems: Observations on their structure, interpretation, and use. Review of Research in Education. Vol. 19 (1993), pg. 451-484.

Burke, W.W. (2014). Organizational change: theory and practice. Sage Publications. 4th edition.

Center on Education Policy (CEP) (2007). Answering the question that matters most: Has student achievement increased since No Child Left Behind? Report issued Center on Education Policy, June 2007.

Childress, S. Higgins, M., Isimaru, A., Takahaski, S. (2010) Managing for results at the New York City Department of Education. In Education Reform in New York City Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Harvard Education Press. Cambridge, MA.

City, E.A., Elmore, R.F., Fiarman, S.E., Teitel, L. (2009). Instructional rounds in education: A network approach to improving teaching and learning.

Coburn, C. E., Honig, M. I., & Stein, M. K. (in press). What is the evidence on districts' use of evidence? In J. Bransford, L. Gomez, D. Lam, & N. Vye (Eds.) Research
and practice: Towards a reconciliation. Cambridge: Harvard Educational Press.

Coburn, C.E., Turner, E.O. (2012a). The Practice of Data Use: An Introduction. American Journal of Education. Vol. 118, No. 2, pg. 99-111.

Coburn, C.E., Turner. E.O. (2012b). Interventions to Promote Data Use: An Introduction. Teachers College Record. Volume 114, November 2012, pg. 1-13.

Cohen, D.K. (1995). What is the system in systemic reform? Educational Researcher. Vol. 24, no. 9. Pg. 11-17, 31.

Cohen, D.K. (1996). Standards-based school reform: Policy, practice, and performance. In Holding Schools Accountable: Performance-Based Reform in Education Helen Ladd (ed). The Brookings Institution: Washington, D.C.

Cohen, D.K., Moffit, S.L. (2009). The ordeal of equality: Did federal regulation fix the schools? Harvard Education Press. Cambridge, MA.

Cohen, D.K., Peurach, D., Glazer, J., Gates, K., Goldin, S. (2014). Improvement by design: The promise of better schools. The University of Chicago Press: Chicago, IL.

Cohen, D.K, Spillane, J.P. (1992). Policy and Practice: The relations between governance and instruction. In G. Grant (ed). Review of Research in Education. 18: 3-49. Washington D.C.: American Educational Research Association.

Cook, S.D.N., Yanow, D. (1993). Culture and Organizational Learning. Journal of
      Management Inquiry. 2: pg. 373.

Daft, R.L., Weick, K.E. (1984). Toward a Model of Organizations as Interpretation
      Systems. The Academy of Management Review. Vol. 9, no. 2, pg. 284-295.

Datnow, A., Park, V. (2009). Large-Scale Reform in the Era of Accountability: The
      System Role in Supporting Data-Driven Decision Making. In A. Hargreaves et al.
      (eds.), Second International Handbook of Educational Change. Springer
      International Handbooks: pg. 209-220.

Datnow, A., Park, V., Wohlstetter, P. (2007). Achieving with data: How high performing
      school systems use data to improve instruction for elementary students. Los
      Angeles: USC Rossier School of Education, Center on Educational Governance.

Dee, T.S., Jacob, B.A. (2010).   The impact of No Child Left Behind on Students,
      Teachers, and Schools. *Brookings Papers on Economic Activity*. Fall 2010.
      Retrieved:
      http://www.brookings.edu/~/media/Projects/BPEA/Fall%202010/2010b_bpea_de
      e.PDF

Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary
      schools: Challenging or reproducing inequality?. The Teachers College Record,
      106(6), 1145-1176.

Elmore, R.F. (2000). Building a new structure for school leadership. The Albert Shaker
      Institute. Washington, D.C.

Elmore, R.F. (2003). The New Accountability: High schools and high-stakes testing.
      RoutledgeFalmer: New York, NY.

Elmore, R. F. (2004a). School Reform from the Inside Out: Policy, practice, and
      performance. Harvard Education Press. Cambridge, MA.

Elmore, R.F. (2004b). Conclusion: The problem of stakes performance-based
      accountability systems. In Redesigning Accountability Systems for Education.
      Susan H. Fuhrman and Richard F. Elmore (eds). Teachers College Press. New
      York, NY.

Elmore, R.F., Abelmann, C.H., Fuhrman, S.H. (1996). The New Accountability in
      State Education Reform: From Process to Performance. In Holding Schools
      Accountable: Performance-Based Reform in Education Helen Ladd (ed). The
      Brookings Institution: Washington, D.C.

Firestone, W.A. (1989). Using Reform: Conceptualizing District Initiative. Educational
      Evaluation and Policy Analysis. Vol. 11, no. 2, pg. 151-164.

Firestone, W.A., González, R.A. (2007). Culture and Process Affecting Data Use in
      Schools. In Pamela A. Moss (ed). Evidence and Decision Making. The
      Yearbook for the National Society for the Study of Education, Part 1. Blackwell
      Publishing: Malden, MA. Pg.132-154.

Firestone, W. A., & Shipps, D. (2005). How do leaders interpret conflicting
      accountabilities to improve student learning. A new agenda for research in
      educational leadership, 81-100.

Fitz, N. (2015). Economic Inequality: It's far worse than you think: The great divide
      between out beliefs, ideals, and reality. *The New Scientific American*. March 31,
      2015.

Foley, E.,Mishook, J., Thompson, J., Kubiak, M., Supovitz, J., & Rhude-Faust, M.

(2008). Beyond test scores: Leading indicators for education. Providence, RI: Brown University, Annenberg Institute for School Reform.

Fuhrman, S.H., Elmore, R.F. (1990). Understanding Local Control in the Wake of State Education Reform. Educational Evaluation and Policy Analysis. Vol.12, no. 1, pg. 82- 96.

Fuhrman, S.H., Goertz, M. E., & Duffy, M.C. (2004). "Slow down, you move too fast": The Politics of Making Changes in High-Stakes Accountability Policies for Students. In Redesigning Accountability Systems for Education. Susan H. Fuhrman and Richard F. Elmore (eds). Teachers College Press. New York, NY.

Fusarelli, L.D. (2002). Tightly Coupled Policy in Loosely Coupled Systems: Institutional Capacity and Organizational Change. Journal of Educational Administration. Volume 40, No. 6., pg. 561-575.

Goertz, M.E., Duffy, M. (2003). Mapping the landscape of high-stakes testing and accountability programs. Theory into Practice. Vol. 42, no. 1. Pg. 4-11.

Hannaway, J., Stanislawski, M. (2005). Flip-flops in school reform: An evolutionary theory of decentralization. In Hess, F. M. *Urban School Reform: Lessons from San Diego*. Harvard Education Press: Cambridge, MA. Pg. 53-70.

Hill, P. T. (2010). Leadership and Governance in New York City School Reform. In Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Education Reform in New York City. Harvard Education Press. Cambridge, MA, pg. 17-32.

Honig, M.I. (2003). Building Policy from Practice: District Central Office Administrators' Roles and Capacity for Implementing Collaborative Education Policy. Educational Administration Quarterly. 39: 292.

Honig, M.I. (2008). District-Central Offices as Learning Organizations: How Socioculutural and Organizational Learning Theories Elaborate District Central Office Administrators' Participation in Teaching and Learning Efforts. American Journal of Education. Vol. 114, no. 4. Pg. 627-664.

Honig, M.I. (2009). What Works in Defining "What Works" in Educational Improvement: Lessons from Education Policy Implementation Research, Directions for Future Research. In Gary Sykes, Barbara Schneider, David Plank (eds). Handbook of Education Policy Research. Routledge: New York, NY. Pg. 333-347.

Honig, M.I. (2006). Building Policy from Practice: Implementation as Organizational Learning. In Meredith I. Honig (ed). New Directions in Education Policy Implementation: Confronting Complexity. State University of New York Press: Albany, NY. Pg. 125-148.

Honig, M.I., Copland, M., Rainey, L., Lorton, J.A., Newton, M. (2010). Central office transformation for districtwide teaching and learning improvement. Center for the study of teaching and policy, University of Washington. Report prepared for Wallace Foundation, April 2013.

Jennings, J. (2012). The Effects of Accountability System Design on Teachers' Use of Test Score Data. Teachers College Record. Volume 114, November 2012, 23

Kane, T. J., Staiger, D.O. (2002). The promise and pitfalls of using imprecise school accountability measures. Journal of Economic Perspectives. Vol. 16, no. 4. Pg. 91-114.

Kaplan, R.S., Norton, D.P. (2007). Using the Balanced Scorecard as a Strategic

Management System. *Harvard Business Review.* July-August Issue. Retrieved: December 2012. https://hbr.org/2007/07/using-the-balanced-scorecard-as-a-strategic-management system

Lortie, D. C. (1975). Schoolteacher: A sociological study. Chicago: University of Chicago Press.

McDonald, J.P., Klein, E., Riordan, M. (2009). Going to scale with New School Designs: Reinventing high school. Teachers College Press: New York, NY.

McLaughlin, M.W. (2006). Implementation Research in Education: Lessons Learned, Lingering Questions and New Opportunities. In Meredith I. Honig (ed). New Directions in Education Policy Implementation: Confronting Complexity. State University of New York Press: Albany, NY. Pg. 209-228.

McLaughlin, M.W., Talbert, J.W. (2002). Reforming Districts. In Amy M. Hightower, Michael S. Knapp, Julie A. Marsh, and Milbrey W. McLaughlin (eds). School Districts and Instructional Renewal. Teachers College Press: New York, NY. Pg. 1173-192.

McLaughlin, M.W., Talbert, J.W. (2003). Reforming Districts: How Districts Support School Reform. Center for the Study of Teaching and Policy University of Washington.

Mintrop, H., Trujillo, T. (2007). The practical relevance of accountability systems for school improvement: A descriptive analysis of California Schools. *Educational Evaluation and Policy Analysis.* Vol. 29, No. 4, pg. 319-352.

Newmann, F.M., King, M.B., Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. American Journal of Education. Vol. 108, no. 4, pg. 259-299.

Newmann, F.M., King, M.B., Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. Harvard Educational Review. 67, 1; pg. 41-65.

New York City Department of Education (2013) What's Next for School Accountability in New York City? Released November 2013.

New York City Department of Education Quality Review Rubric (2015). Retrieved: http://schools.nyc.gov/NR/rdonlyres/8C11A001-7E78-469D-996F-B0C3703CEA81/0/201314QualityReviewRubric.pdf May 21, 2015

O'Day, J. A. (2002). Complexity, accountability, and school improvement. In Redesigning Accountability Systems for Education. Susan H. Fuhrman and Richard F. Elmore (eds). Teachers College Press. New York, NY.

O'Day, J., Bitter, C. (2010). Improving Instruction in New York City Schools: An Evolving Strategy. In Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Education Reform in New York City. Harvard Education Press. Cambridge, MA.

Oakes, J. (1989). What educational indicators? The case for assessing the school context. Educational evaluation and policy analysis. Vol. 11, no. 2. pg. 181-199.

OERI (Office of Educational Research and Improvement) (1988). Creating Responsible and Responsive Accountability Systems. Report from OERI State Accountability Study Group. U.S. Department of Education. Washington, D.C.

Orton, D.J., Weick, K.E. (1990). Loosely-coupled Systems: A Reconceptualization. The Academy of Management Review. Vol. 15, No. 2, pg. 203-233.

Panero, N.S., Talbert, J.E. (2013). Strategic inquiry: Starting small for big results in education. Harvard Education Press: Cambridge, MA.

Personal phone interview Rajeev Bajaj, February 20, 2013.

Peurach, D. (2011). Seeing complexity in public education: Problems, possibilities and success for all. Oxford University Press.

Peurach, D., Glazer, J., Gates, K. (2004). Supporting instructional improvement: teacher learning in comprehensive school reform. Washington, D.C., National Clearinghouse on Comprehensive School Reform.

Polikoff, M., McEachin, A., Wrabel, S., and Duque, M. (2013). The waive of the future: School accountability in the waiver era. Paper presented at the Association for Education Finance and Policy Annual Conference. March, 2013. New Orleans, LA. Retrieved:
http://www.aefpweb.org/sites/default/files/webform/AEFP_Waivers_Final.pdf

Popper, M., Lipschitz, R. (1998). Organizational Learning Mechanisms: A structural and cultural approach to organizational learning. *Journal of Applied Behavioral Science*. 34:161.

Porter, A.C. (1991). Creating a system of school process indicators. Educational evaluation and policy analysis. Vol. 13, no. 1 (Spring). pg. 13-29.

Rich, M. (2015). Reviewing the federal education law, senator seeks more local control. *The New York Times*. January, 13, 2015.

Rosenberg, S. (2012). Organizing schools for quality in education: Individualistic and systemic approaches to teacher quality. Dissertation: University of Michigan

Rowan, B. (2001).The ecology of school improvement. Journal of Educational Change, 3, 283-314.

Rowan, B. (2002). Rationality and Reality in Organizational Management: Using the Coupling Metaphor to Understand Educational (and other) Organizations – a Concluding Comment. Journal of Educational Administration. 40, 6: pg. 604.

Saez, E. (2013). Striking it richer: Evolution of top incomes in the United States. *Pathways Magazine*. Stanford Center for the Study of Poverty and Inequality. Winter (2008) original publication, pg. 6-7. Retrieved: http://eml.berkeley.edu//~saez/saez-UStopincomes-2012.pdf May 21, 2015.

Scott, W.R. (2000). Institutional change and healthcare organizations: From professional dominance to managed care. University of Chicago Press.

Selden, R. W. (1990). Developing educational indicators: A state-national perspective. *International Journal of Educational Research*. Vol. 14, issue 4. pg. 383-393.

Shipps, D. (2012). Empowered or beleaguered? Principals' accountability under New York City's diverse provider regime. *Education Policy Analysis Archives,* 20 (1).

Smith, M.S., O'Day, J. (1990). Systemic school reform. In Politics of Education Association Yearbook. Pg. 233-267.

Sparks, D., Malkus, N. (2013). First-year undergraduate remedial coursetaking. Institute of Education Sciences (IES). National Center for Education Statistics: Statistics in Brief: January 2013.

Spillane, J.P., Parise, L.M., Sherer, J.Z. (2011). Organizational Routines as Coupling Mechanisms: Policy, School Administration, and the Technical Core. American Educational Research Journal. Vol. 48, No. 3, pg. 586-619.

Sherer, J.Z., Spillane, J.P. (2011). Constancy and Change in Work Practices in Schools: The Role of Organizational Routines. Teachers College Record. Volume 113, Number 3; March 2011, pg. 611-657.

Stecher, B., Hamilton, L., Gonzalez, G. (2003). Working smarter to leave no child behind: Practical insights for school leaders. RAND White Paper ISBN: 0-8330-3477-4.

Supovitz, J. (2010). Knowledge-based organizational learning and for instructional improvement. A. Hargreaves et. al. (eds) *Second International Handbook of Educational Change*. Springer International Handbooks of Education 23, pg. 707-723.

Taylor, J., Stecher, B., O'Day, J., Naftel, S., LeFloch, K.C. (2010). State and local implementation of the No Child Left Behind Act Volume IX – Accountability Under NCLB: Final Report. National Longitudinal Study of No Child Left Behind and the Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind. Prepared for the U.S. Department of Education, Policy and Program Services.

Weick, K. E., Roberts, K. (1993). Collective Mind in Organizations: Heedful Interrelating on Flight Decks. Administrative Science Quarterly. Vol. 38., No. 3, pp. 357-381.

Zavadsky, H. (2012). School Turnarounds: The essential role of districts. HEPG: Cambridge, MA.

CHAPTER III

MANAGING FOR IMPROVEMENT: THE DESIGNS OF THE PROGRESS REPORTS

AND QUALITY REVIEWS IN NEW YORK CITY

The key is that unless there is accountability, we will never get the right system. As long as there are no consequences if kids or adults don't perform, as long as the discussion is not about education and student outcomes, then we're playing a game as to who has the power.

- Albert Shanker, 1993

Very simply, I seek a school system that ensures a quality education for every student, where results are not a function of student demographics, but are the product of a system that values achievement by all students…I seek a system that holds managers responsible for the success and failures of individual units based upon the results that parents have the right to expect, measurable, relevant educational achievements by all our children.

- Mayor Michael Bloomberg, 2002

Since the 1960's, state and school districts across the U.S. have been publishing school report cards. Seen as critical tools for promoting accountability for schools and local districts, school report cards have been used to publicize data about school effectiveness to parents, policy makers, educators, and other stakeholders. The logic was that when key decision makers were armed with the data, they would be equipped to see where and which schools were succeeding, and where there was room for improvement.

The practice of publishing report cards dates back to federal Elementary and Secondary Act of 1965, which was amended in 2001 and 2008. Under Title I, Part A of the law, States and local educational agencies (LEAs) are required to prepare and disseminate "an annual report card" that must include information about public schools related to "student achievement, accountability, and teacher quality, as well as any other material the SEA or LEA deems relevant (DOE, 2013).[1] The original goal for Title I, which remains true today, was to improve educational equity by providing better educational opportunities for low income students with the help of more federal funds for school districts serving higher percentages of poor students. The report cards were meant

to help the public know which schools were effective, and in turn, whether they were equitable.

Conventional report cards offer indicators about student performance and include information like aggregate and disaggregate percentages of students scoring at each proficiency level on a state's standardized assessment; this information reflects whether a school is making progress towards the state's annual goals for Annual Yearly Progress. The reports often describe quantitative data about schools including two-year trends in student achievement, high school graduation rates, and teacher qualifications. Lastly, most versions specify which schools have been targeted for interventions due to persistent low-performance.

Report cards, or school score cards as they are called in some states like Michigan, use what many have referred as lagging indicators to spotlight schools' past performances. Borrowing from business management, indicators are referred to as lagging when they characterize an organization's past performance. By comparison, leading indicators would predict future performance (Foley et al, 2008). Education has largely used lagging indicators because they were feasible with the technology that was available and it was believed that this kind of information could be a critical resource to parents, policy makers, and educators alike. The logic was that armed with the indicators, the different stakeholders would seek optimal outcomes and make decisions accordingly. Once people were making informed decisions and pursuing optimal outcomes, the quality of education for all students would rise. But the score cards were limited and offered little guidance for how to raise quality or broaden access to quality options.

In 2002, the New York City Department, under Mayor Michael Bloomberg's control, developed a radically different set of accountability tools that were meant to offer more balanced information for decision-making, which would help fix the schools and increase quality options for all students.

Hardly a conventional politician, Mayor Bloomberg proposed a different approach to school accountability. In 2007, under the Mayor's control, the New York City Department of Education (NYC DOE) instituted Progress Reports and Quality Reviews to evaluate schools and report school quality information thought to be useful to parents,

educators, community leaders, and central office. The tools included lagging and leading indicators thought to offer guidance and focus attention. The lagging indicators reflected in the Progress Reports included measures of students' progress on state tests and the learning environment in schools. The addition of *leading* indicators in the Quality Reviews, which included measures of schools' curriculum, pedagogy, and structures for improvement (i.e., goals and action plans) were intended to signal when mid-year corrections were necessary and allowed stakeholders to track whether schools were developing capabilities and managing assets in ways that were likely to produce better student achievement. Combining leading and lagging indicators, the NYC DOE was trying to create a performance management system that could balance accountability for past results and promote continuous improvement toward future outcomes (Childress et al, 2010).

Mayor Bloomberg's and Chancellor Klein's strategy was based on the premise that schools and the NYC system would deliver improved educational and equitable outcomes through better management techniques. The mayor had inherited a system that was described by his predecessor, Mayor Rudolph Giuliani as, "just plain terrible – it makes no sense, and the end result of it is that, if this were a business system, it would be in bankruptcy" (Guiliani in Fullan, 2014, pg. 21). NYC's old Board of Education's reputation was that it was a compliance-driven, conservative bureaucracy in which innovation was rare and corruption rampant (Klein, 2014). The Mayor and Chancellor sought to transform the system – it's culture, structure, and outcomes.

Towards these ends, in 2002, Mayor Bloomberg fought for and won formal authority to control New York City's school system. His first move was to replace the Board of Education with a thirteen-member Panel for Educational Policy (PEP) of which he appointed eight of the seats. By gaining all of the power, Mayor Bloomberg eliminated points of friction that had stymied past reforms and squashed power squabbles so that people could start to focus on improving the delivery of quality educational experiences to all students (Gyurko and Henig, 2010).

Under a seemingly centralized strategy, central office was made responsible for providing conditions and tools that would permit strong management practices in schools; yet, this was combined with a decentralized approach that granted autonomy to

all principals and schools were made responsible for managing for instructional improvement. Central office developed several accountability tools and instruments, including the Progress Reports and Quality Reviews, to guide schools in their attempts to meet their new responsibilities and measure the effectiveness of their attempts. Central office had to develop and manage these tools as supports for schools that were charged with improving education for all students – especially the lowest performing students. For example, one of the original architects explained, "I think it was really clear that we knew that we needed to motivate people to be attentive to how their students were doing and that that had to be focused on every single student and we wanted to instill a sense of deep responsibility for all of the kids." The DOE used the Progress Reports to communicate this message. However, central was also aware, as this architect explained, that, "secondly then, we wanted to create the resources and tools that would enable educators who came to that view and said, okay 'I feel responsible and I see failure happening,' provide them with a set of tools that they could use in a strong way and we realized this would take a lot of work too, to provide those tools to them."

The tools, the Progress Reports and Quality Reviews, served as a compass of sorts among the reforms that Mayor Bloomberg and Chancellor Klein implemented. The leaders sought innovation and improved performance – neither of which was possible under the former system where thirty-two local superintendents held more sway than central office in their separate fiefs. Mayor Bloomberg reduced the power of the superintendents significantly and downsized the encumbered bureaucracy to create a dynamic central office – something akin to a start-up company equipped with talent and resources to revolutionize an industry. This new central office was responsible for designing and managing the novel accountability tools, which evolved over time. Despite the bold and swift change occurring in the larger environment, the accountability tools were meant to serve as guideposts and cultural artifacts during the system's transformation to a more centrally coherent and openly innovative system where everyone was responsible for increasing the quality of education available to all 1.1 million students in New York City public schools.

To enable people, mainly principals, to use the new accountability tools and manage their schools for improvement, the designers of the reforms also put in place

other building blocks of human and civic capacity. Mayor Bloomberg and Chancellor Klein solicited support from influential foundations like Carnegie Foundation of New York, Open Society Institute, Bill and Melinda Gates Foundation, as well as individual philanthropists who also supported reforms based on management principles and market-based ideas (Gyurko and Henig, 2010). Some of the philanthropy money was used to create the New York City Leadership Academy to support the development of a talent pipeline of leaders equipped to handle the financial, human and instructional management that school leaders were expected to perform as empowered school leaders.

After eliminating many layers of bureaucracy, the Mayor and Chancellor also established a district-designed marketplace for school support organizations. Empowered principals, who controlled their budgets, self-affiliated with a network, referred to as the Children First Networks. The networks provided administrative support and professional development for their partner schools; the networks were intended to help schools, especially those that could not figure out how, to meet new expectations for improved and innovative school performance and as such became critical scaffolds for learning. Yet, the networks were also new and needed to develop their own capabilities one step ahead of the schools they worked with. These unique organizational entities were placed in an even more complicated position because the Mayor and Chancellor created a bi-level accountability structure that held the schools and networks accountable for results. Given this, the accountability tools became paramount for schools and the networks.

The overlay of this new structure, which included different governance, new support providers, new tools, and steep consequences for failure on top of an existing culture that possessed many norms that were potentially incompatible in the new system created all manner of complications. With this essay, I do not aim to judge the success of addressing the complications; rather, I argue that the case New York City's system reform provides evidence that converting ideas into practice depends on learning to do so. A clear plan and steely commitment are necessary, but reformers seeking to replicate NYC's system-reforms risk disappointment, if they see principled plans and new structures as sufficient.

In the sections that follow, I provide details about the development of the Progress Reports and Quality Reviews and their evolutions amid the intentionally disruptive

revolution that was Mayor Bloomberg's and Chancellor Joel Klein's Children First initiatives. To tell this story, I borrow a framework from other scholars who sought to "explore, understand and critically analyze the increasing ubiquitous work of large-scale, sustainable, systematic educational reform" (Peurach, 2011, pg. 16). While this framework has been used by researchers to explore comprehensive school designs enacted by organizations like Success For All and America's Choice that attempted to operate within, as well as, outside of the formal system of U.S. public education (Peurach, 2011), I am using this framework to explore a single-site case in New York City's designs for its unconventional school accountability tools.

Analytical Framework

I start with the assumption that to describe the designs of New York City's school accountability tools, I need to do more than simply describe or define their pieces. I begin with this assumption because simple descriptions would suggest that these tools could have been designed anywhere – Chicago, Los Angeles, or Grand Rapids. Offering a basic description of the tools, while interesting because of how rare they are in the U.S., would miss the contribution of the complexity of interactions among the environment in which these tools were created – the systems they were part of, and the emerging changes that they sought to influence. The interactions among these things provided resources and challenges that influenced the ultimate pieces that would make up the tools, as well as, their evolution.

I draw from a long line of influential education policy and implementation scholars who argued for greater appreciation for the complexity involved in large-scale reform initiatives (i.e., Berman & McLaughlin; 1978; McDonnell & Elmore, 1987; Cohen & Hill, 2001; Schon & McDonald, 1998; O'Day, 2002; Honig, 2006; McLaughlin, 2006; Peurach, 2011). These scholars offered valuable advice for using a systems view for research over alternative approaches that looked at singular policy domains or categorical reforms like curriculum development or teacher preparation; a systems view seeks to understand the inherent interrelatedness of how people, policy, and place interact as intentional, planned change is planned for, and how it occurs as reformers attempt to change systems (McLaughlin, 2006; Honig, 2006).

Part of the rationale for considering, confronting, or seeing complexity in

education research is the very fact that interdependencies and relationships exist in an educational system (Peurach, 2011). When a researcher does not account for these, she may find something significant, while another researcher may conclude the opposite –and both may be 'right.' Taken together, the findings leave the field confused about what we actually know about a particular policy or its implementation (Honig, 2006; Peurach, 2011). Rather, than consider what works, it is better to ask, what works under what conditions as conditions influence outcomes (Honig, 2006). Given this, subsequent research that attempts to explain the outcomes should consider conditions and their intricacies (Honig, 2006; McLaughlin, 1991).

A second part of the rationale for valuing complexity has to do with the nature of problems and solutions in education. Peurach (2011) draws on the work of Herbert Simon to describe the usual ways of thinking about problems in complex systems. Simon's suggestion, according to Peurach, was that complex systems may resemble a hierarchy of nested "boxes-within-boxes" that are "nearly independent in the short term and only weakly interdependent in the longer term" (Peurach, 2011, pg. 16). Given the near-independence of the elements, those seeking reform or change, may design improvements by isolating an element, seeking to understand it and targeting it for improvement. As Peurach described, Simon referred to this as the "empty world hypothesis." Using an empty world hypothesis, which is conventional practice in education, problems are treated as though they might be addressed reasonably with a targeted intervention that is designed and disseminated in a straightforward, or linear fashion (Peurach, 2011). The subsequent result is that the field ends up collecting research that speak of dismal failures and pages filled with suggestions for next time that offer little help for practitioners today (McLaughlin, 2006).

In contrast, Peurach (2011) calls for a "full world hypothesis" about education. The full world hypothesis assumes that "many things are strongly connected to many other things" (pg. 17). A key difference between empty and full world hypotheses can be seen, if we imagined a broken wrist-watch. This is clearly a problem. To solve the problem, according to the empty world hypothesis, we could take apart the watch, disassemble it and isolate and target the broken piece. Once the part has been identified, it can be removed and repaired or replaced. The repair allows this complex system, the

watch, to return to working order; problem solved. A full world hypothesis does not work the same way because the full world hypothesis assumes that all of the parts are interrelated – isolating the problem is more difficult because of the interrelatedness. Rather than a wrist-watch, we might think of a cloud. You cannot take apart a cloud to fix it. Rather, the cloud is an emergent system composed of many interrelated and interdependent parts. While this is not an exact analogy, it demonstrates an important distinction when it comes to solving problems in complex systems. A full world hypothesis suggests interdependence of the parts that make up the system; solving problems in this kind of system, then, requires one to see the system (Peurach, 2011)[2].

Using this view to understand education problems, according to Peurach (2011) permits reformers to "see" complexity in education, "…a world full of at-risk students; a world full of complex parts, problems, solutions, and challenges, all in dense, interdependent, networked relationships; and a world full of individuals, organizations, and groups working in interaction to apprehend, confront, and reform these many parts and their dense, interdependent relationships" (pg. 17).

It is important to understand that designs for reform do not fall on 'empty worlds' or what Joseph McDonald referred to "blank slates." Rather, policies are arguments for particular action that interact with existing, lived realities of those for whom the policies are intended to help, motivate, or direct (Cohen & Moffitt, 2009; McDonald et. al., 2014). Rarely is policy starting at day 0; neither does a new policy mean that we get to start over. Rather, policies enter into worlds where there are rich histories including valued lessons that were learned in the past and kept; policies interact with existing world-views, networks, and relationships – these among other things provide resources for policy, or their existence may signal a need for additional, or different resources that policy designs must account for. A slate filled with text offers many opportunities, but may also introduce challenges for those attempting to manage the work involved in designing and implementing ambitious and novel designs in a particular environment over time (McDonald et al, 2014; Cohen, Peurach, Glazer, Gates, & Goldin, 2014).

Lastly, there is an important point about complexity in education that often gets

---

[2] Source: Credit goes to David Brooks for discussion of wrist watch and cloud as distinguishing between types of systems. Brooks describes the cloud as an example of an emergent system. Brooks, D. Tools for Thinking. *The New York Times.* Published March 28, 2011.

overlooked in education research – namely, that complexity presents a challenge for how researchers attend to variation. "As organizational theorists have long shown, complexity in its various forms – including variation – can serve as a stimulus for innovation and improvement especially given the diverse and sometimes unpredictable circumstances under which educational leaders routinely operate (Honig, 2006, pg. 22). However, education research has struggled to distinguish variation that arises from artful solutions and the kind of variation that stems from incongruous patchwork policy. Too often, studies are limited and restrict investigations to find common, standard, or average results in order to make generalizable claims that are good for taking a policy to scale. As a result, variation in implementation is often characterized as failure and attributed to inadequate designs that overlook practical considerations, a hijacked process, poor understanding by or weak capabilities of those responsible for implementing the policy, but the problem may actually have been a misspecification of the design by the researchers.

It is also possible that it is not a misunderstanding at all; rather, scholars draw conclusions based on a general acceptance that the research they conducted was the research that *could* be done. Understanding complexity is an enormous and expensive undertaking. When scholars privilege manageable research projects and limit the scope of a study (which is completely reasonable), they may settle for less thorough investigations.

McDonald et al. (2013) addressed this point when they argued, "sensitive to the complexities of real reform in practice, an evaluator must instead consider which goals to track and which outcomes to count. We think such an evaluator should also try to herd diffuse intentions - not only as espoused in original plans, but also as wired into designs, and implicated in actions - into a coherent theory of action" (pg. 581). Yet, this is not a road often traveled as there are few maps for this kind of research. When researchers consider implementation, they seek to generate ideas about how meaning and doing relate in the context of a particular reform. We need better knowledge about how to understand the gaps between what a policy means and what gets done; broader conceptions of the complexity in designs could be a useful starting place (Yanow, 1997; Schön & McDonald, 1998; Peurach, 2011; McDonald et al, 2014).

Given this view, I drew on the work of scholars who sought to understand how organizations working outside of the education system sought to create programs that could improve as many of America's weakest schools as possible. More specifically, I used the framework that David Cohen and his colleagues Don Peurach, Joshua Glazer, and Simona Goldin described in their recent text, *Improvement by Design*. The formula for this framework was developed over more than a decade with the Study of Instructional Improvement and was the basis for two dissertations, a book, and several articles. The authors refer to this framework as a way to consider "improvement by design as a strategy for large-scale, practice focused, school improvement" (pg. 21).

The authors used four intersecting puzzles to tell a comparative story about Success for All, Accelerated Schools Project and America's Choice. The four puzzles include the design puzzle, the implementation puzzle, the improvement puzzle, and the sustainability puzzle. The scholars choose the analogy of puzzles because, as they and others have suggested, reform efforts do not unfold in orderly progression from one phase of work to the other; rather, reform is less rationale.

To describe a reform from the point of the interveners – not just policy makers who sit above schools, but to tell the story of those who held specific ideas about schools' problems and what the solutions might and how they then tried to build solutions requires an organizing scheme to make sense of the many "tangled" decisions and actions that "often unfold in overlapping and nonsequential ways" (Cohen et al, 2014, pg. 26). The authors settled on the analogy of puzzles as a framework through which they tell the story of how the interveners managed different aspects of activities unfolding over time and how those actions attempted to "untangle and manage complex interactions among schools, designs, and intervening organizations, and to do so in environments that are at least intermittently turbulent" (pg. 26). In this essay, I used their conception of a design puzzle to tell the story of New York City's blueprints for their Progress Reports and Quality Reviews.

The Design Puzzle

According to Cohen, Peurach and their colleagues, the design puzzle "centers on efforts to construct plans and blueprints for developing coherence and capabilities among

schools, designs, intervening organizations, and environments, with the goal of supporting improvement in practice and achievement" (2014, pg. 22). This puzzle addresses the following domains: the *schools* that were to be the targets of the reform, the *designs* that interveners constructed to support the schools' improvement, the *organizations* that interveners created to develop and field the designs, and the *environments* in which schools, designs, and organizations operated (2014, pg. 22).

In the sections that follow, I use these four domains: schools, designs, organization, and environment to tell the story of New York City's Progress Reports and Quality Reviews as they were conceived of and constructed during the period of time when Mayor Bloomberg was in charge of the city's schools. I begin by relaying what I saw to be Mayor Bloomberg's and Chancellor Klein's core strategy and how that attended to the city's schools and what it meant for the central office organization, and how the environment intersected with these.

Under Mayor Bloomberg's charge, the New York City Department of Education was trying to change radically, in a relatively short period time, how the country's largest school district was managed. As one of the original architects shared, "You can't tell a kindergartener that you're going to take five years to change the schools – that's half his life; you will have wasted his chances. You can't take forever with incremental, institutional change. You have to be willing to make big, bold changes that disrupt stuff. So, the challenge is knowing how to balance 'listening and learning tours' with hitting the go button."

My aim with using the design puzzle framework as an organizing scheme is to attempt to capture the dynamic, ambitious, and novel nature of the changes that NYC's leaders sought when they decided to hold schools accountable for performances measured by the Progress Reports and Quality Reviews. Where my depiction departs slightly from the authors' original delineation of four puzzles – the design puzzle, the implementation puzzle, the improvement puzzle and the sustainability puzzle – is the relationship between implementation and design. I did not set out to study implementation, but the story of New York City's designs for their accountability tools were intended to evolve, and as such, activities that happened during implementation affected the designs.

According to Cohen and colleagues, the implementation puzzle "centers on what happens when plans and blueprints are put into motion, including the unanticipated and dysfunctional interactions that arise among schools, designs, intervening organizations and environments" (pg. 23). This suggests that the implementation puzzle considers what happens as designs are put in motion and reformers find "it very difficult to realize their core ambitions" (pg. 61) and the implementation puzzle reflects how they "manage interactions within and among those four domains (schools, designs, organizations and environment) over time in order to realize their potential" (pg. 61). The implementation puzzle is concerned with the problems and unanticipated challenges that arise as interveners go through implementation.

The design puzzle deals with how reformers tried "to structure conditions" strategically among the four domains "in order to increase the potential for effectiveness" (Cohen et al, 2014, pg. 61). The designs in NYC involved planned change. Poole (2004) suggests the difference between planned and unplanned change: Unplanned change rarely occurs as a direct result of strategic policy decisions. It may or may not come from intentional efforts. By contrast, planned change is a conscious effort to improve a situation and has an end goal as a reference point (Poole, 2004). Planned change suggests a level of advanced choreograph that happens to invite changes. The designs in NYC were specifically intended to evolve and in the descriptions below, I will refer to the evolution of the tools, which took place during implementation over time, but I categorize this as part of the design because the architects were conscious about developing ways for the tools to adapt in order to make sure there was an adequate fit with the schools and those meant to use them - as the core strategy depended on people in the system using the tools as critical resources in their improvement efforts.

Methods and Data

In this essay, I present a description of the blueprints for New York City's school accountability tools that relies on the design puzzle framework to trace the assumptions, resources, and capabilities that went into building a new kind of school accountability model. I arrived at these blueprints by conducting a study that sought to answer three research questions: (1) What is the design of the Progress Reports and Quality Reviews?

(2) How do Networks use the tools in the service of school improvement? and (3) What influence does the environment in New York City have on the design and use of the Progress Reports and Quality Reviews? To answer these questions, I spent a year shadowing one of the Cluster Leaders, which gave me direct access to the eleven networks and their partner schools (n=304). For this study, I used a case study design (Yin, 1994) taking as the cases a) the design of Progress Reports and Quality Reviews in New York City and b) Network Leaders'[3] use of the tools in their work supporting schools. In this essay, I report on the case of the design of the tools; in a second essay, I address the network leaders' use of the tools.

Data

Over the course of year, from June 2013-July 2014, I conducted sixty-eight semi-structured interviews with current and former Department of Education (DOE) staff including those on the Cluster team, network teams in the Cluster, principals and central office staff. The interviews lasted between thirty and ninety minutes, and I met with at least seven participants more than once (this is not reflected in the total number of interviews). Several of the people I interviewed served in many different roles during their careers with the DOE; in the description that follows, I refer to these participants by a title that closely reflects the role they were in at the time of our interview. There were several respondents who were no longer DOE employees when I met with them; I refer to these participants as former staff, or original architect, when applicable.

In addition to interviews, I observed more than ninety hours of NYC DOE, Cluster, and Network meetings and collected upwards of eighty documents. I used snowball and purposive sampling to gain access to relevant people, meetings, and internal documents; my prior experiences as a teacher in a NYC high school and as an intern with

---

[3] The governance structure of New York City's school support system was completely redesigned under Mayor Bloomberg's control. By 2010, there were 6 Clusters in NYC (reduced to five in 2012). Each Cluster contains at least 10 Networks; each Network provides instructional and operational support for 30 schools. Network support is not geographically based; rather, each principal selects the Network with which to partner. Each Network has a Network Leader, who is responsible for managing relationships and support for schools in the Network. For more on the changes in NYC governance structure, see Hill, P. T. (2010). Leadership and Governance in New York City School Reform. In O'Day, J, Bitter, C.; Gomez, L. (eds). Education Reform in New York City. Harvard Education Press. Cambridge, MA, pg. 17-32.

the NYC's Department of Education helped me make valuable connections and secure access. Every one I contacted, except one person, agreed to speak with me.

The limitations of this study include that it is an oversized project conducted by a lone researcher over the course of one year. I relied, in many ways, on retrospective accounts of a reform that spanned more than a decade in the country's largest school district. Retrospective accounts are valuable, but limited especially given two truths: the first is that by the time I arrived, some of the dust had settled. When Mayor Bloomberg and Chancellor Klein initially introduced the disruptive innovations in governance, accountability and change management, those working in the system were required to make swift and abrupt shifts from how they used to understand the world. One of the limitations I faced was being able to capture what was once new, but had become taken for granted. One senior official in the DOE expressed this point when he was describing the Quality Review, he said, "I don't think there's much argument around here anymore," but there used to be. Over time, the starkness of an experience becomes dimmed as experience teaches us that everything worked out in the end.

Another limitation of this study is its ability to fully reflect that these reforms were specifically designed to evolve. One architect explained, "…we went through a bunch of iterations…we changed it every year and in a very self-conscious planned way that we talked about it and we changed it every year." This same architect highlighted how such an effort poses a challenge for research. He explained, "When we created my center, one of our goals, which we have not fulfilled by any means, is to think about what it means to conduct research on an organization that is self-consciously changing. So, it's not fair to that organization to study things as if there's an X and you're going to see the effect of X because their goal is to set X in motion and then react to it and change it, but it also does put a huge burden on researchers to figure out how to do that." With this study, I tried to find a working solution for that burden.

Lastly, the timing of this study offered an interesting opportunity given that during the middle of the study, New York City had a mayoral election. In the months before the Mayor left office, the Department of Education was both reflective and anticipatory. Several scholars and think tanks offered assessments of Mayor Bloomberg's legacy, which included reviews of the accountability tools that were created

under his watch; similarly, scholars like Aaron Pallas from Teachers College offered suggestions for Mayor Bill deBlasio, who took office in January 2014. The timing of my data collection proved to be generative, but complicated, as the new administration's transition in January 2014 did not go smoothly. It was a precarious time for many of the participants in my study; I remained sensitive and open to their realities, which allowed me to continue learning and collecting data despite the seemingly turbulent times. The constant comparisons between Mayor Bloomberg's 'old' accountability and Mayor deBlasio's 'new' accountability system provided me a an interesting lens through which I was able to consider substantive differences between an evolutionary change in a design and a completely new tool.

Analysis

Data analysis began with data collection and was ongoing and comparative, which allowed me to refine my ideas about the designs (Miles & Huberman, 1984; Weiss, 1994; Maxwell, 2005; Corbin & Strauss, 2008). All interviews were recorded and transcribed; I coded transcriptions using Dedoose, a qualitative analytic software. Initial codes were based on the research questions and protocols and provided basic descriptions of the designs of the Progress Reports and Quality Reviews. I used analytic memos to refine and deepen my understanding of patterns as they related to the design puzzle framework (Corbin & Strauss, 2008). Based on these emerging ideas, I continued to develop and refine codes to reflect the complexity and interactions among the tools, schools, central office, which I treated as the organization managing the tools and the broader environment. I used multiple sources of data including the coded interviews, observations and numerous artifacts to triangulate and guard against threats to validity (Maxwell, 2005). When possible, I used member checks and consulted with the Cluster leader I shadowed, network team members, and original architects to confirm my general sense of the designs.

      In the following section, I offer a basic overview of Mayor Bloomberg's and Chancellor Klein's core strategy for school accountability and provide background information on how these tools fit within a larger reform movement referred to as the Children First Initiative. I then describe the blueprints for the Progress Reports and

Quality Reviews and present how the strategy attended to designs, schools, central office, and the broader environment. I want to highlight that this is not an evaluation of the designs; I do not intend to declare that parts of the design were good, while others were bad. The point of this essay is to tell the story of the designs, not to judge them.

There are certainly lessons to be learned from NYC's bold experiment, but a recipe for replicating the city's strategy is not among them. The descriptions in this essay offer foundational knowledge of the complexity involved in converting a strategy for system reform from a set of arguments for fixing educational problems into a set of meaningful artifacts that were trying to move all of NYC's classrooms located in seventeen-hundred schools across all five of its boroughs.

To arrive at the designs for the Progress Reports and Quality Reviews, the NYC DOE had to strike a balance among a set of tensions that inherently exist in school accountability that by its very nature has many goals. Such models, as the DOE acknowledged, "focus attention on – and delineate expectations for desired student outcomes. They provide metrics that that the district and its stakeholders can use to monitor school performance. They serve as a learning tool for educators highlighting success areas or areas in need of improvement. They draw attention to schools that are performing well relative to their peers, providing models of good practice, and provide a basis for intervention or closure when a school is performing poorly" (Corcoran & Pai, 2013).

The rationale for the strategy resided in the artifacts, press releases, presentation decks, and in conversations among DOE staff. During one meeting in April 2014, I was sitting around a table with several support and central office staff who were discussing how to help "weak schools" develop an "instructional focus" which was a requirement outlined in the Quality Review rubric. One veteran network team member threw up her hands in frustration and asked, "Well how do you do it? How do you move a system?" Without a pause, a colleague from her cluster team answered, "You don't have to move a system. You have to move a classroom. I can hear Joel saying it now. He was insistent: 'You don't have to move a system. Parents don't send their kids to a system; they send them to a classroom. They don't even send their kids to a school, in most cases, they send their child to a classroom with one teacher for 180 days a year.' That's what you

have to focus on – moving that one classroom." In NYC, system-wide reform was seen as that simple. Yet, as history has shown, it has also proven very rare. This is what Mayor Bloomberg and Chancellor Klein set out to do and was the main reason why they built a new kind of accountability model for the country's largest, and some might say, most complex and contentious school system.

Core Strategy

Mayor Bloomberg's and Chancellor Klein's strategy for moving classrooms and the system was based on management principles. The presumption was that you could improve educational outcomes for all students, if the adults in the system were more results-oriented. According to a report that the Chancellor's office distributed to the public, "Our goal is to focus everything we do on the only outcome that matters: student success" (NYC DOE, 2008).

Mayor Bloomberg and Chancellor Klein believed that system level reforms were necessary because the old system had "…a wobbly history of overbureaucratization, overfragmentation, patronage politics, and stark inequalities" and did not put students first (Gyurko and Henig, 2010, pg. 93). What allowed the system to persist like this for so long was that people blamed things like poverty and constrained control over the things that affected student outcomes; there was a pervasive view that the schools were doing the best they could given challenging circumstances. The Mayor and Chancellor held a different belief and were motivated to make disruptive changes because they believed that children's chances for success in life were dependent on the quality of the education they received in school (Klein, 2014). According to the Chancellor, a school system should be designed, operated, and managed to deliver the best quality education to every child – no matter what. The Mayor and Chancellor sought a service delivery model that did not oversimplify instruction, nor did it ignore the challenges of teaching in low-income, high poverty neighborhoods; rather, it was based on a belief that adults working in the system – whether they worked in a school, classroom, or central office cubicle, were responsible for solving any problems that prevented a child from receiving the best educational experience that would enable her to live a productive life after school. According to the designers, *every* student in New York City deserved the highest

quality education and not just the best possible given circumstances, but the best without exceptions or excuses. To transform the system they inherited, Mayor Bloomberg and Chancellor Klein, with their team of leaders, developed a strategy to accomplish four related tasks: (1) make problems explicit, (2) assign responsibility, (3) balance educator and performance management perspectives, and (4) evolve with continuous improvement practices.

Make problems explicit

The intention under the Bloomberg-Klein strategy was to move away from bureaucratic, top-down control of schools because it was believed that a system that served diverse students needed diverse schools. Those closest to classrooms were in the best position to influence students' learning and create innovative solutions for serving the diverse populations of students. The goal was to "create not a great system, but a system of great schools" (Hemphill, 2010; Klein, 2014). Applying a business-minded lens, great schools were believed to be learning organizations that could adapt to meet the needs of students. The strategy sought to convert the old compliance-oriented cultures that treated collective joint-work as a direct threat to individual autonomy into learning organizations.

Scholars, many outside of education, have described learning organizations, but offered few instructions for how to build them. Learning organizations have adaptive capabilities that help the organization, or company, remain competitive in a dynamic market. Fulfilling a "Darwinian imperative" these organizations adapt and seek to improve continually to stay ahead of the field (Supovitz, 2010). According to Senge (1990), learning organizations do more than simply survive; rather, they possess "adaptive learning" and "generative learning" that "enhances our ability to create" (pg. 14). Further, Senge suggested that learning organizations are distinguished as places "where people continually expand their capacity to create the results they truly desire" (Senge, 1990, pg. 3).

The descriptions suppose that these kinds of organizations have command of their separate departments and are able to coordinate work practices towards some shared goal. Following this, these organizations focus on internal core practices while keeping an eye on the external environment in order to answer to customers and competitors by offering

innovative solutions. Learning organizations are similar to what Daft and Weick (1984) referred to as information systems; these types of organizations deliberately scan their environments, collect information, and use that information to take collective action in response. Edmonson and Moingeon (1998) similarly define learning organizations as those in which the "members actively use data to guide behavior in such a way as to promote the ongoing adaptation of the organization" (pg. 28). According to Senge (1990) a way to help organizations become more flexible and adaptive is to "discover how to tap people's commitment and capacities to learn at all levels" (pg. 4).

Learning organizations are able to adapt to solve problems, but problems are only evident when everyone who is contributing to the organization is aware of their shared goal. They know how to correct and detect performance related problems. However, in education detection of error, or success, has been determined based on students' test scores. Using this as a gauge, one could imagine, as explained to be one of the executives in NYC DOE that problems are only recognized when students' test scores drop, or remain persistently low. Using these kinds of gauges, people do not investigate problems in the schools because they appear to be problems with the students. This kind of system suggested that, "the way you got be a good school depended a lot on who entered your school." Another architect explained, that according to this kind of logic, if you wanted better schools, you had to get better students.

The strategy in NYC sought to change this thinking and suggested that schools needed to manage the quality of education and solve organizational and technical problems in order to provide better quality educational opportunities, which would be visible in improved student outcomes. Outcomes were based on the value provided by teachers and principals - not simply how well any students did on a test. To accept this view, many educators had to comprehend that school quality was more than just a function the pedigree of students; rather, the new mental model framed school quality in terms of the contributions of teachers and principals.

Part of the strategy was to name the characteristics of quality schools – the tasks, processes, and behaviors performed in schools that were believed to make a school good at moving student learning, and therefore, if a school possessed these traits, it could be considered a high quality school that influenced students' positive growth. For example,

one senior director at the DOE explained, "It's [the Quality Review] meant to outline what are the qualities of an effective school community. Like what are the categories of things you need to keep track of, and think about, and plan for. What does it look like when those things are in place and interacting with one another? And so, yeah, the rubric itself, it should, it's envisioned like a map to a high quality school."

A key part of the strategy was to use the accountability tools to serve as "maps to high quality" schools. These maps would serve as orienting guides for networks that were charged with helping schools develop capabilities and to direct schools – especially schools unskilled with developing internal capabilities to embed the types of behaviors specified by the DOE and less equipped to perform the tasks thought to increase student achievement.

Stacey Childress and her colleagues sought to understand New York City's approach for using accountability to promote the development of capabilities to improve student achievement. They report that the original architects sought to "attempt to strike the right balance between instilling accountability for past results and encouraging improvement toward future outcomes" (Childress et al, 2010, pg. 92). Further, they suggest that the goal of the designs was to help schools "develop into organizations in which professionals were constantly learning with one another about how to solve performance problems" (pg. 92). They quoted Jim Liebman, NYC's first Chief Accountability Office, who described how the designers viewed the purpose of the tools:

Accountability isn't entirely or evenly mainly about incentives. It's about capacity building, which to me means adult learning based on self- and team evaluation of what's working and what's not, and knowledge management, meaning spreading what works from one student or school to another. If we want the lever of accountability to be as powerful as possible, we have to provide ways for schools to build their capacity to be relatively self-sufficient in evaluating themselves every day and in solving their unique performance problems and, when necessary, in asking for the specific help they need. This will never work, if the central bureaucracy behaves as if it has all the answers. Our role is to help professionals in schools ask better questions so that they can craft customized answers based on their evaluation of their performance problems (in Childress et

al, 2010, pg. 92).

In this way, the designs for the tools suggested that when principals used the tools, they would acquire organizational learning capabilities. The tools were prompting these behaviors by evaluating leaders – not just on the answers they came up with, but also on the questions they asked. The tools required principals to show their work and calculations.

This kind of problem-solving in education is what Judith Warren-Little refers to as a "strong technical culture" which was an extension of what Ball and Cohen (1999) refer to "learning in and from practice." What these authors suggest is that teachers can improve their pedagogy, or practice, when they engage in "deliberate, mindful investigation of student teaching and learning" (Little, 2007, pg. 218). Little suggests that a "vision of systematic investigation of practice, like the image of evidence-based decision making, promises certain thoughtful and informed consideration of what constitutes a crucial question and what evidence of practice might bear on that question…to exploit daily experience for purposes of professional learning would also seem to suggest certain norms of professional discourse among colleagues – for example, what constitutes an acceptable, useful, or valued story of practice" (pg. 218). The strategy for the Progress Reports and Quality Reviews were based on an idea that these tools could require interactions that promote greater professional discourse among colleagues, and provide more evidence of practice that could inform what was talked within those interactions – ultimately, strengthening the professional culture in schools.

For example, one of the initial architects and former Chief Accountability Officer explained,

> And then to put it altogether and to get this deeply embedded in the schools, we created the inquiry teams and inquiry teams and senior achievement facilitators were designed to have in every school experts among the teachers who would be using all of these tools in a structured way to identify and focus on the actual kids who were, as we put it, outside the sphere of success at that school and would have a real structure for identifying those kids, studying them, observing them in a variety of ways including low-inference observation where it's just pure observation and using all of the data that was available and other data that they

could generate by adopting certain little assessments that you'd use, hypothesize the problem where instruction was failing. We're very clear, it's not a problem with the kids, this is a problem with what you're doing. It's working with some kids but it's not with others so that the instruction is failing. What can you do to improve the instruction? And then run that experiment, we didn't use the word experiment, but run that effort and set goals for yourself like, what would success look like? And measure periodically, it could be every week, every few weeks, every month, whether you're succeeding, and then if not, figure out why, go through the process again a little bit to figure out why and adjust to do that…

What the designers were seeking were ways to promote the development of norms for learning in and from practice within schools.

Mehta (2013) argued, "For the most part, schools are not places which could be characterized as centers of inquiry or places with continuous improvement. Scholars of teachers have described it as a profession heavy on 'presentism, individualism, and conservatism'" (pg. 281). The designers in NYC had to contend with identities and attitudes that ran counter to this view of accountability – one where accountability is seen as a learning tool and not just a gotcha when educators are out of compliance.

The strategy in NYC attempted to disrupt many people's long held views about what it takes to improve quality education in schools, and the mental models about quality schools, in general. For example, the senior director who referred to the Quality Review as a map explained:

I mean, I think one of the biggest barriers is one that I already talked about, which was like how do people think about school quality. Do they think, 'Oh, a school is just running like the trains are running – on time and like kids seem to be well known. That's a good school. And yet, going into ten classrooms, five of the classrooms are not places where students are asked to do any thinking. So, I think a barrier of such breakthrough has been redefining what a high quality school is and how it is deeply attached to the quality of classroom instruction and the quality of student work that is coming out of those classrooms – that indicate that kids are engaged in critical thinking. I think that was a barrier because that frankly, I think that there were pockets of places where there really low

expectations for what a good school was and while having your kids safe and well known are obviously really important and foundational to being able to do academic work, they are not the end point. I think that was probably the biggest barrier was the mental model of what a good school is.

The strategy, in many ways, tried to confront this, by attaching stakes to the Quality Review and Progress Reports – and integrating the two. By making everyone accountable for the tools, people in the system were required to pay attention to problems that happened when they strayed off course, or headed down the completely wrong path.

Acknowledging the fact that many principals were going to need targeted support to operate and meet the new expectations for problem-solving and improved performance, the Mayor and Chancellor created agencies of support for principals that were referred to as the Children First Networks. These external support providers that principals self-selected were also held accountable for schools' results. This arrangement created strong incentives for networks to mobilize resources and provide relevant and effective professional development for schools around the accountability tools.

Mayor Bloomberg and Chancellor Klein were neither naïve nor unaware of the colossal challenges they faced in trying to convert the New York City Board of Education into a performance-results-oriented Department of Education where everyone in the system was focused on getting every student ready for college and successful career. The premise these leaders operated from was that "The pre-2002 school system does what it was built to do: make stable jobs, accommodate the demands of interest groups and comply with state laws. It can do those things without providing effective schools for kids. We intended to rebuild the system to put children and learning first," as explained by one of Klein's closest associated (in Hill, 2010, pg. 4). Yet, they also knew from their experiences leading businesses and management literature that changes in structures would only go so far; they also sought to change the culture. They targeted culture in three ways: accountability, leadership, and governance.

According to DOE's report (2009), *Children First: A common-sense plan to create great schools for all New York City children,* the Chancellor and his team sought to change the system in three ways:

(1) From a culture of excuses, where educators too often blame students and their

families for low performance, to a culture of accountability, where adults take responsibility for ensuring that *all* [sic] children, regardless of their circumstances, learn and achieve; (2) From a culture of compliance, where educators waste too much time doing paperwork and following one-size-fits-all directions from administrators, to a culture of achievement, where the central focus is on results and doing whatever it takes to help each student learn; (3) From a culture of top-down bureaucracy, where central and regional offices make most decisions, to a culture of individual great schools, where principals and their teams design the programs that their particular students need to succeed (pg. 4).

It is a difficult thing to know whether structures drive culture, or vice versa. Mayor Bloomberg targeted both simultaneously.

Assign responsibility

The Klein-Bloomberg strategy was based predominantly on management principles. From management of the system, to management in schools, the designers sought to change how decisions were made and work was performed across the system, which also included changing relationships and responsibility for performing tasks. The designers did not presume to know how to teach or even what should be taught; rather, they stacked the leadership team with experts who had strong instructional records. It was these few leaders who would set the terms for many of the reforms and serve to constrain Klein and other lawyers on the team (Hill, 2010). The DOE recommended certain math and literacy curriculums – for which they provided professional development opportunities, but in line with empowerment, principals were allowed to select their curriculums. Central office required that the principals justify their selection to ensure that people were making informed decisions and not just picking convenient curriculums, or ones not aligned to the Common Core Standards, which the city adopted in 2009. Like most things under the Bloomberg-Klein strategy, schools needed to justify decisions as being related to results, and central office was responsible for providing the conditions to support decision-making.

    After leaving his role as Chancellor, Joel Klein issued a manifesto for fixing

public schools that was published in the *Washington Post*. In the manifesto, he explained that a performance-based culture comes from adults being able to make choices and then held accountable for those choices. The role of central office was to create the conditions – fair and open – that would allow principals and parents to "choose something better for their kids" (Klein et al, 2010). Giving parents, principals, and teachers options drove the version of decentralization that Klein and Bloomberg championed.

The plan for decentralization was based largely on the work of William Ouchi, a professor from UCLA's Anderson School of Management (a business school) (Childress et al, 2010). Ouchi is well known for his Theory Z, which was explained in his book, *Theory Z: How American Business Can Meet the Japanese Challenge*. Ouchi's Theory Z is an expansion and counter to other theories of motivation and what drives workers to perform well (see Douglas McGregor's (1960) Theory X and Y in *The Human Side of the Enterprise*). At risk of oversimplification, Ouchi proposed a management style under which workers are trusted to do their jobs to their utmost ability and the role of management is to support them to do that. Writing in the 1980s at the time when American corporate culture and industry (especially the automotive industry) were facing fierce economic competition from the Japanese, Ouchi who studied Japanese firms suggested that "productivity gains are the direct result of an involved workforce, that coordinating and organizing people is the most important task of management, and that productivity is a matter of sound social organization rather than hard work" (Bisesi, 1981, pg. 82). Ouchi proposed a management philosophy, or a new paradigm, that said skillful management that focused more on coordinating and integrating social elements like norms and routines across activities between stages in processes of an organizations work (i.e., design, manufacture, and delivery) would lead to better outcomes than rational bureaucracies that focused on specialization and technical solutions (Ouchi, 1981).

Ouchi, a scholar of large organizations, was perplexed why large companies like McDonalds or Toyota, two companies he studied extensively, could deliver high-quality service routinely across diverse local contexts, not some of the time, but all of time, and why school districts could not figure out how to do the same (Ouchi, 2013). "McDonald's built its reputation by making everyone of its stores deliver high quality customer service to every customer every time – not by delivering high quality service

sometimes to some customers" (Ouchi, 2013, pg. 3).

Ouchi observed that there were pockets of excellence in every school district, and yet, the districts were managed such that some students were given good service, where as many of the students who most needed good service were given sub-standard service. He saw 'service' and 'education' as interchangeable, and he positioned districts as the 'company' in charge. Why was it that big companies like Eli Lilly, an organization that has remained prominent for more than a century, with its research facilities located in seven countries could figure out how to deliver reliable, consistent, high-quality service, but school districts could not?

After more than a decade of research and visiting more than six hundred schools across the U.S., Ouchi concluded that there was a deficiency of excellent schools because school districts were not managed in ways to support the transformation of schools, or expansion of excellent options. Rather than top-down, bureaucratic control, he argued that districts should be decentralized.

Ouchi's version of decentralization was different than the first wave of decentralization that occurred in response to America's education crisis during the 1980s. The crisis was marked by the release of the National Commission on Excellence in Education's 1983 report, *A Nation at Risk* that concluded our country's foundation was "eroded by a rising tide of mediocrity." Purkey and Smith (1985) explained that this first wave of decentralization was based on an assumption that education would improve, if decision-making and accountability moved closer to the child and classroom. During the decades that followed, decentralization policies were matched with the development of national standards for what students should know and increasing emphasis on test-based accountability, but twenty years later, Ouchi observed that pockets of excellence and widespread inequality remained. He was bothered about the persistence of inequality and dismayed that education research seemed to suggest that these patterns were mostly attributed to factors outside of schools' control.

Ouchi proposed a different version of decentralization that focused on five pillars: (1) school choice, (2) development of effective principals, (3) accountability, and (4) weighted student formula budgeting (Ouchi, 2013). This form assumed that giving authority to those close to schools to be decision makers, and holding them accountable,

could only work, if supported by the district central office. According to Ouchi (2003), "It would be more accurate to say that change should be *initiated* bottom-up and *supported* top-down. That is to say that the central office should not be imposing new ways to run a school top-down on principals. Instead, a central office is one that gives principals the freedom to experiment for themselves" (pg. 251).

Influenced by Ouchi's ideas, and other leading reform theorists like Charles Sabel and Michael Barber, as well as his own experiences leading the global media conglomerate, Bertelsmann, Chancellor Klein and his team developed the NYC's accountability-autonomy. The autonomy-accountability granted all principals autonomy to make decisions about budgets, staffing, professional development, curriculum, interim assessments, and support. Chancellor Klein's opinion was that, "school principals need to have autonomy in staffing and running their schools – and to be held accountable for results…real educational quality can't be created by top-down fiat; it has to take root at the school level" (Lehigh, 2014). [4] However, the designers did not view strict accountability as the main lever of improvement; rather, accountability was part of a productive tension with organizational learning. The locus of control was given to principals so that they could become the lead problem-solvers in schools. With principals leading inquiry and problem-solving, it was thought that inquiry would lead to organizational learning and the schools would become learning organizations. Yet, it was the responsibility of central office to provide the conditions through which this was possible.

During a presentation to The Organisation for Economic Co-Operation and Development (OECD) in 2010, one of the lead architects explained that the NYC DOE created an infrastructure for "Building K-12 Learning Organizations" that required the central office to fulfill four tasks: empower, enable, evaluate and enforce.   Referred to as the 4E Framework, these tasks were performed by different offices within the DOE, but were seen as necessary and related to accomplish the larger goal of creating a system of learning organizations.

The 4E framework served to organize the various streams of activity that central office had to manage and perform. For example, "enable" involved activities that

"frequently assess student learning strengths and weaknesses, provide data to educators, and train educators to work in teams to diagnose and cure instructional failure." The presenter suggested that "evaluate" activities included, "rate schools and educators based on student learning, given student challenge (lagging indicators), qualitatively review schools and educators based on strategic use of available tools (leading)."

One of the original members of the Office of Accountability shared with me, "We talked about the four Es – empowerment to empower the educators to have the ability to do this. We weren't in charge of that. But then, evaluation, enablement, and enforcing consequences was really part of evaluation. And that was the theory." According to this architect and others, the theory was that DOE's evaluation of schools' performances should also serve to enable and enforce consequences, which was a fairly ambitious and unprecedented pursuit. In 2007, there were no good models in the U.S. from which NYC could borrow, or learn. Instead, the NYC DOE attempted to learn from monitoring their efforts (Hill, 2010).

Balance educator and performance management views

Early on, trying to determine what to evaluate involved debate among senior leadership. Despite the evidence in research and common skepticism about test-based accountability, there was still a contingent in the Bloomberg-Klein administration that was not convinced that test scores alone were insufficient. This group imagined that the Progress Report need only contain test-based metrics as this was seen as the measure that best demonstrated the quality of "products" that schools produced - namely, knowledgeable students. A senior official in the DOE explained that, "the chancellor's [Joel Klein] idea was overly simplistic. He wanted to hold schools accountable for student performance simply on the basis of standardized test scores, which he had much greater confidence in than most of us who'd actually worked in the system - who understood what they were cable of and what the limitations of them were." As this quote highlights, there were camps among the leadership - one camp was seen as educational outsiders; the other belongs to those coming from within the system. The distinct identities were recognizable by the beliefs held by members about what accountability needed to look like and how it would predictably behave. How these camps came to consensus shaped much of the

early designs for the Progress Reports and Quality Reviews would eventually shape implementation of what would become the accountability tools.

In one camp were mostly people new to education who had come to the DOE from successful careers in other fields like business and law. Those from this camp believed that, if targets were set accordingly, successful schools would be able to meet the targets and those schools that could not would be closed. This view was reminiscent, and in some cases drawn from something like management by objectives, or performance management. Performance management is a term borrowed from management literature that was just beginning to be used by transplants - leaders new to education reform around the early 2000's. The general idea behind this perspective is that when managers (central administrators) set clear, measurable goals for the organization (school system) it drives outcomes because doing so directs everyone in the organization towards the goal. When there are clear, measurable targets, actions of individuals in the organization (schools in the system) can be monitored and assessed by how well they contribute to larger organization's goal. A bumper sticker for this camp might read, 'what gets measured gets done.'

There was another camp of people who had been working in the system for a long time - on the fringes. This camp's roots go back in time to the late 1980s with a set of reforms that were championed by Tony Alvarado. Alvarado, as Chancellor of NYC schools in 1983, founded a city-wide sub-district called the Alternative High School District. The Alternative School District promoted small-schools, innovative instruction, and a new kind of school management free from central office red tape.

Much of this history is outlined by Joe McDonald in *Autonomy and Accountability in New York City School Reform*. McDonald wrote, "Alvarado's intention in launching the district was to foster creative insubordination within an immense bureaucracy practiced at stifling creativity" (2009, pg. 7). What once was considered insubordination would become central's expectation under Klein, but it would be referred to as school autonomy.

Many of the schools that were part of Alvarado's Alternative District would gain recognition far from NYC including those that were part of the Coalition of Essential Schools and Center for Educational Innovation (CEI). The Alternative District was tied

to the Community School District's legacy and Deborah Meier's Central Park East. New Visions for Public Schools was incubated and developed by people participating in the Alternative District; these leaders would become part of the educator camp under Klein (McDonald, 2009). While the Alternative High School effort struggled to spread across the city, it did create generations of like-minded people who shared dispositions towards instructional novelty, organizational conditions, school leadership and networked support free from bureaucratic control.

This second group, the educator camp's ideologies would form some of the basis for the Empowerment Zone under Klein's Children First campaign. Those working in the Empowerment Zone were skeptical of anything that bolstered bureaucracy or overlooked the role of context. According to this group, if leadership wanted different outcomes, then capacity had to be built for teachers and principals to work differently - emphasizing that teachers needed opportunities learn how to improve the quality of instruction in classrooms and that principals needed opportunities to learn how to be instructional leaders equipped to support teachers and manage conditions for instructional improvement. People in this camp knew, mostly from experience working in and with NYC schools, that building an infrastructure capable of moving perspectives and practice took more than setting targets and putting pressure on outcomes; rather, they saw the need for frameworks to direct attention and scaffolds for supporting the diverse learning needs among educators, school leaders, and central office staff.

While my description of the two camps is useful, in some ways, it may also oversimplify the diversity of ideas and contributions that would eventually produce the Progress Reports and Quality Reviews. What description delivers is a window into a world where debate was not just permitted, it was encouraged. "By negotiation with each other and lots of proposals about ideas and lots of vetting of those ideas and then of course the chancellor would make many of the decisions, some of them ended up being the mayor's decision, out of that came this combination of things that we did, " as explained by one of the members of the original design team.

He went on to explain, that it was, "kind of Dewey and kind of public problem solving…but it was a mixture." The strategy was a mixture – a mixture of business related concepts, but also education related concepts about improving schools. The

problem is that within the system (and most education reform debates), this mixture is not well received, or a natural blend. Rather, it takes disrupting strong beliefs about the nature of problems and solutions. Yet, in NYC, the architects tried to create a mix of business and education related concepts for improving schools.

Part of the rationale for creating a mix was that the DOE was really trying to use the accountability system to accomplish a set of goals. One goal was to use the tools to develop continuous improvement practices in schools; the other was to use the tools to signal to central office, which schools were making improvements and which were not so that central could make some decisions about what to do with those schools. One former DOE executive explained the effort to manage the different goals:

> There were [competing views] – I think in two ways. One is that there's this sort
>
> of kind of comprehensive accountability idea that I described where it's basically
>
> all an effort to generate this kind of continuous improvement process – is one vision
>
> what the accountability system is about. Another vision of what the accountability
>
> system is about…is it's a way to make personnel decisions, basically, portfolio
>
> decisions. The two aren't mutually exclusive. Even if you have a great sort of
>
> continuous improvement scheme set up, there may be circumstances where someone
>
> who is not a good fit and they need to go, or where a school is just so broken that it
>
> needs a really fresh change, but there's, I think there's, it's possible to emphasize
>
> both. And there are people in the department who emphasize one, and there are
>
> people in the department who emphasized the other.

One of the architects, drawing on the "Pragamatists," explained that blending the two views to find the right balance between business and education presented a challenge because the two camps do not necessarily conceive of quality performance. Therefore, they draw different conclusions about what constitutes evidence of quality. In business, quality is often based on target goals and outcomes. The architect referred to this as "deduction." He explained, "…deduction, which is the target people want – they want just want to know, 'Did you meet your target?'" According to him, the other side, the educators draw conclusions about quality using "induction." The educators, according to the architect (and education scholars who put forth the view of teaching as a artistic craft, or style), "induction, which is what teachers like to think about craft, 'we just know

74

intuitively'" which makes measurable evidence across teachers and schools hard to come by.   Both views are limited. "Well, deduction won't work because we don't know what the target should be and induction won't work because good teachers will do it and the other teachers won't and we'll never know what the good teachers did and it typically works with the best kids who are a already well-performing," explained the architect. Neither deduction, nor induction would help central office or principals monitor performance and manage for improvement.

Instead, the DOE's strategy was to use "abduction." According to the architect, "abduction…is just this way that you structure inductive reasoning as much as possible to use every bit of knowledge that you already have and to make it as explicit as possible and use as much data as you can even if you're not doing a rigorous kind of experimental study." The point was to require principals, and teachers, to reflect on and "acknowledge what you know and don't know, use what you know, try to figure out and solve what you don't know."

In 2007, using an abductive approach, the DOE required all schools to have an inquiry team, which consisted of a small team of staff – teachers and administrators who collectively worked to target low student performance in their schools. Upon targeting a focus group, they were supposed to investigate reasons for low performance, propose and design an intervention, and learn from the experiment with the expectation that the lessons learned would have possible implications for school policy. The DOE provided tools and facilitators to schools to help them conduct and have ownership over an internal inquiry process.

By 2010, the Quality Review rubric included indicators that evaluated schools on the use of inquiry teams[5], but the expectation was that the practice had expanded from one team in each school to teams of teachers within the school who were expected to be working collaboratively on problems of practice. The most recent rubric included Indicator 4.2 that evaluated schools on their capability to "Engage in structured professional collaborations on teams using an inquiry approach that promotes shared leadership and focuses on improved student learning." The inquiry teams became a way to combine a focus on targets and encourage busy educators to focus on improving

---

[5] See Panero & Talbert (2013) Strategic inquiry: Starting small for big results in education.

practice through inquiry.

The debates between the two camps were a continuous source of tension for the designers, but in many ways the competing views reflected "public debate about these things," as expressed by one former director. He explained, "I think the public debate is really a set of caricatures – you have the 'fire your way,' 'the excellence crowd,' and then you have 'those who defend traditional educational prerogatives at all costs,' and there's not a lot of attention paid to, at least two more moderate middle between those positions, where actually a lot of people tilt…but share a lot of common ideas." He further explained that within the wider debates, "it's much less common to see in the public debate people really advocate for a continuous improvement process."

In his view, and others that I spoke with, this is really what NYC's strategy was trying to advocate for – a continuous improvement process at the school, and system level, but it took some learning on the part of the architects to know how to approximate that in an education setting (as complex as NYC's). One former chief executive in the DOE expressed,

> I think there definitely was more of a commitment over time that Joel [Chancellor Klein] came to understand and appreciate of learning from all of – explicit improvement in what's happening through explicit adult learning where adults work with each other, figure out and can actually explain what they did and make that available to other educators who, of course, would then have to customize it, to some extent, to their situation and their kids…that would be more explicit than just having it be some kind of hidden market forces or hidden empowering of great educators.

Balancing the educator and performance-management views took time and a commitment to learn from the designs, which reflected the final part of Bloomberg-Klein strategy.

Evolve through continuous improvement practices

An essential element in the initial blueprint was that the tools would evolve over time. One former executive explained, "We sought to change the bureaucratic model, but the reforms were specifically designed to change. Evolution and adaptation were built into the model. That's part of instilling continuous improvement. Like Toyota, we had to be self-reflexive. If we wanted people in schools to focus on continuous improvement, we

too had to respond to feedback. The reforms were based on continuous improvement principles."

To create a fair, comprehensive, and equitable accountability system that could realize marked improvement, the original architects included a process to revise the Progress Reports and Quality Reviews each year that involved monitoring and learning from the tools as they were used. For example, one former architect, who lead the office responsible for the Progress Reviews talked about how his office would conduct "listening tours" and collect feedback on the Progress Reports in order to revise them to better fit their intended use. He commented,

> I guess in that sense – the kind of collaborative process of identifying and then working to solve problems – a level of detail that we were trying to encourage in the schools was also being replicated at the systemic policy level with the Progress Report, and Doug Knecht can tell you more about the Quality Review, but my understanding is that they had a group – pretty similar process – for revising the Quality Review.

Similarly, people who managed the Quality Review discussed the "four or five month process every year" that they used to collect feedback from all constituents – principals, networks and superintendents about the Quality Review process. Drawing from feedback collected during the tours, evidence in education literature (i.e., Judith Warren-Little, Richard Elmore, Charlotte Danielson, and Mike Schmoker, etc.), and evaluations conducted by the DOE's internal Research and Policy Support Group (RPSG), the Quality Review team made changes to the rubric to in an effort to "prioritize what we thought was most important," according to a former executive in the DOE; the point was to use continuous improvement principles on the designs for the tools. In this way, the DOE was mirroring, at the system level, the processes and dispositions for managing for improvement what they wanted school leaders to embed in routine practice in schools.

Designs

The NYC DOE intended the Progress Reports and Quality Reviews to create coherence and stability for its transformation of the system. These tools went through various evolutions, but the point was always to "measure capacity" explained one of the original

architects. In the sections that follow, I explain what the Progress Reports and Quality Reviews were by focusing on their essential elements.

The Progress Reports evaluated schools annually based on four areas: School environment, student performance, student progress and additional credit. Schools were awarded points for each of these areas and then compared to schools that served similar populations of students and the city to determine each school's overall grade, A-F (See Appendix A). The Progress Reports offered a historical account of how a school performed in the past. Despite the challenges the DOE faced with fairness, fit with the state's accountability system and reliability, these reports quickly became the way people characterized schools (Childress et al, 2010). The Quality Reviews were a two-day site visit by an experienced educator who used a rubric created by the DOE to evaluate the school based on three domains: The instructional core, culture of learning and structures for improvement (See Appendix B). The evaluator would observe classrooms and meet with teachers, parents, students, and administrators to get an on-the-ground view of the school (DOE NYC, 2013). Principals would conduct a self-evaluation of the school. Evaluators would use the self-evaluation and their observations to discuss with principals the areas of celebration and areas in need of further focus. Schools would be given a rating – Well Developed, Proficient, Developing, or Underdeveloped and this would appear on the front of the school's Progress Reports.

What is essential to understand about New York City's Quality Reviews is that New York City was one of the first districts in the United States to require all of its schools to undergo a Quality Review as a form of accountability. Given this, New York City had to define its Quality Review; it relied on help from Cambridge Associates. In 2009, the DOE managed the Quality Review independently and created its own version of the Quality Review rubric, which served as a different kind of management tool for schools and central office.

The architects of the tools were aware that simply creating the tools would not guarantee their use. To help schools use the tools and the information they conveyed in the reports, the DOE created networks of support that were intended to help schools with all manner of operational and instructional matters – including knowing how to use the Progress Reports and Quality Reviews as they related to each school's unique situation,

which many schools were not able to determine on their own. The networks became an instrumental agency in helping spread understanding and help those in schools embrace the tools as useful, relevant parts of the regular happenings in schools. Central office had to devise ways to inform networks and principals about the tools and encourage the use of these tools by the support providers in their relationships with schools. The DOE took on the task of creating the tools, as well as, creating infrastructure to support communication, professional development and monitoring of use among networks and schools. For a central office that had been fairly weak and exerted little influence over system-wide efforts in the past (as most reforms and development were handled by local superintendents), this creation of the tools and the support infrastructure were enormously ambitious and challenging tasks.

Achieving accuracy through comprehensive measures on the Progress Reports
The Progress Reports were distinctly different from other score cards used in states and districts across the U.S in two important ways. The first is that they aimed to be accurate depictions of school quality – measuring school quality by more than student test scores. The second is that they attempted to create a fair accountability model – holding schools accountable for student progress taking into account different challenges schools faced in moving their students. They sought to accomplish this by evaluating schools through peer comparisons.

The first noteworthy feature of the Progress Reports is the comprehensive nature of the metrics used to calculate the grades listed on each school's Progress Report. Several of the original architects of the reports explained that these reports had to be comprised of more than test scores, in part, because the Reports were part of an agreement made with principals: More autonomy for more accountability. Given this, the tools could not simply report how well students were doing on tests; people did not trust the tests to be good gauges. Test scores could be inaccurate, inconsistent and removed from what people working in schools cared about. If you reject the messenger and do not respect him, you certainly will not invest in what he tells you to do - especially if what he is telling you to do is really hard and requires you to do something new, different, or unproven to lead to better results.   The architects knew they needed a better messenger

than a list that ranked schools based on students' test scores.

Instead, the thinking was that it was important to measure "the ability of a school to help students progress toward the eventual goal of earning a Regents diploma. The measures should focus on the capacities students develop as a result of attending the school, not the capacities they bring with them on the first day" (NYC Educator Guide, 2013). Some in the department referred to this as being like a value-added model - asking what did the school contribute to student learning? "We felt that was the only fair and productive way and, in fact, defines what value added assessment is all about." Relying on test scores alone would not have been a fair determinant, nor would it been taken very seriously by principals. If not test scores alone, then what?

Since 2006, when the reports were first implemented, schools were evaluated on the quality of contributions made to students' learning outcomes in four main areas: (I) Student progress, (II) student achievement, (III) school environment and (IV) closing the achievement gap. More than half of a school's grade is based upon on whether students make progress towards a diploma. The DOE designed the Progress Reports to measure "the ability of a school to help students progress toward the eventual goal of earning a Regents diploma. The measures focus on the capacities students develop as a result of attending the school – not the capacities they bring with them on the first day" (NYC DOE Educator Guide, HS, 2013). (See Appendix C: Progress Report Overview).

For progress in particular, the DOE placed particular emphasis on the one-third of students who entered at the lowest-performance levels and gauged how well the school helped them make progress towards graduation. One senior official in the DOE, who had also been a principal when the Progress Reports were first introduced, explained, "It became a way to insure that schools were feeling accountable for every child, that's what I think was the most powerful about some of those data sets that came out, rather than looking at, well I'm just going to push my 2s to here and like there was that kind of talk for a long time, I want to say probably since the late 90s, maybe the early 2000s, there was that talk around, well let's just group them [students], put our 1s here, our 2s here, 3s here, 4s here, like we could move this amount of 3s to 4. I think what the progress report did was be able to nuance and feel accountable for every student, not just trying to push 2s to 3s."

In its Educator's Guide to the Progress Reports, the DOE explains, "The Report is designed to encourage principals and teachers to accelerate academic achievement toward the goal of career and college readiness for all students. By tracking student academic progress, identifying steps to improve each students' learning, planning a course of action to achieve that improvement, and revising the course of action needed to ensure progress, our schools can ensure that every student leaves school prepared for the next step in his or her education" (NYC DOE, 2013). To account for different targets and desired learning outcomes for students in high, middle and elementary schools, the DOE created separate versions of the Progress Reports.

When the reports were first published in 2007, there were only two types of reports: one for elementary and middle schools and one for high schools. Based on feedback from stakeholders, over the next six years, the DOE expanded the types of reports to six to account for the unique realities of transfer schools, early childhood programs, young adult centers, and District 75 (special education) schools. The DOE published all six annually, and released guides every year to outline changes to each type.

For elementary and middle schools, 85% of the overall letter grade was based on state exams. Some observers, including Dan Koretz, critiqued this method highlighting that state's tests were designed to be point-in-time measures and not year-to-year growth indicators; opponents argued that the measures were flawed because state test scores fluctuated from year to year making them less than accurate gauges of what any one school might contribute to its students' learning (Kelleher, 2014).

One of the original architects explained that the DOE was aware of the potential risks of placing so much emphasis on state exam scores, but decided to use them for middle and elementary schools, in part, because those were "the measures of student outcomes that were available at the time. For middle and elementary progress reports, the state test results for grades three through eight in English and math were available." Further, he explained, "there was also a science exam for fourth and eighth graders, and a social studies exam, but for various reasons those are not considered as high quality." In 2010, in response to concerns like these, as well as evaluations conducted by DOE's Research Policy Support Group (RPSG), an internal, but independent, research center that conducted data analysis and evaluations for the DOE and feedback from parents,

principals, and educators, the DOE changed the metrics for the elementary and middle schools to include growth percentiles.

A student's growth percentile is calculated by comparing her growth to all of the students who started at the same proficiency level the year before. An example that the DOE uses in training presentations for principals, network leaders, and superintendents goes like this: In third grade, Sally earned a 2.84 proficiency rating on the State Common Core math test; as a fourth grader, she scored 3.29 proficiency rating on the State Common Core math test. 84% of the students who, like Sally, scored 2.84 in third grade scored 3.29 or below in 4th grade, which puts Sally in the 84 growth percentile. Another way to consider this is that a student in the 50th growth percentile made 50% more growth than the students who started the same place.

The DOE adjusted growth percentiles based on student demographics to account for errors that occur in approximating relative growth; it is a faulty premise to suggest that all students start in the same place. For instance, there are weighted adjustments for students with different types of special education status, English language learners, and percentage of students qualifying for free lunch under Title I (a proxy for SES).

Individual student growth percentiles are used to calculate two different indicators on a school's Progress Report: median growth percentiles for all students in the school and the median growth percentile for the students in the bottom third of students based on prior year's test results. According to Kelleher (2014), the Independent Budget Office released a report in 2012 claiming that "this change in methodology increased the year to year stability of the student progress sub-score, indicating it improved the Progress Reports' ability to capture 'systematic rather than spurious' differences between schools" (pg. 23).

For high schools, the overall grade on the Progress Reports was based on a broader set of measures. According to one senior official at the DOE, "They were more successful at being balanced at the high school level where there were numerous internal indicators to balance the external indicators." Decisions about which metrics to include were based upon what was available, as well as, what made sense given the desire to encourage principals and teachers to focus on student progress and elements known to predict desirable achievement outcomes. For example, the original designers "included

credit accumulation based on historical data that showed a strong correlation between the number of credits a student had in specific semesters (especially ninth and tenth grade) and the probability of graduating on time" (Childress, 2010, pg. 97).

In its initial form, the Progress Reports included a section called "Closing the Achievement Gap." In this section, schools could earn up to sixteen extra-credit points for achieving "exemplary outcomes among high-need students" (NYC DOE Educator Guide HS, 2013, pg. 5). Among those students that were considered "high-needs" were students with disabilities, English language learners, black and Hispanic males in the lowest-third citywide. Schools could earn extra credit for moving special education students to the least restrictive environments. Schools could only earn bonus points; this section was not used to deduct points from schools' overall grades.

Over time, the DOE expanded the measures to reflect "local and national priorities, particularly a focus on college and career readiness" (DOE, 2013, pg. 4). In 2010, the DOE added a college and career readiness section to the high school Progress Report. This section included five sub-indicators: College prep course index, four-year and six-year readiness indexes, postsecondary enrollment by 6 month after high school, and postsecondary enrollment 18 months after high school. In the first year of adding these metrics, the DOE reported the results, but did not count them towards accountability; in its second year of using these metrics, the DOE did use them to calculate schools' grades.

When the Progress Reports were first implemented in 2007, they included internal indicators from the non-academic data that was available (I.e., attendance). The DOE also attempted to measure school climate by included results from the Learning Survey, was distributed to parents, teachers, and students and asked questions in four areas: Academic expectations, communication, engagement, and safety and respect. Drawing on survey development at the University of Chicago's Consortium on Chicago School Research (CCSR), the survey was intended to assess community perspective on a school's culture, instruction, and systems for improvement and provide school leaders with insight into school's strengths and areas for improvement. The DOE distributes the survey annually to all parents, students, and teachers in the city - making it one of the largest surveys in the country (second to the U.S. Census) (Nathanson et al, 2013).

In 2008, the DOE partnered with the Research Alliance for New York City Schools

for assistance with analyzing and evaluating the survey and its results. One of the findings reported by the Alliance in 2013 was that a key strength of the surveys was their robust response rates from teachers and students (83% and 78% respectively). The researchers concluded that these increasing rates "offer confidence that the survey results reflect the broader population" (pg. 4). In addition, the survey items "provide statistically reliable indicators of the DOE's four reporting categories" (pg. 6).

However, the relationship between survey results and test scores and graduation rates was not consistently clear. Relating an anecdotal view of the surveys, one DOE official explained,

"There has been controversy especially in the field about the use of the survey information. People feel that because it's subjective and it's people's views about schools that it's not a good way to evaluate the quality of the schools. Do parents really know the quality of the schools? They are not in the building every day. Teachers could use the survey as an opportunity to get back at their principal if they are upset with their principal. Our kids really taking this seriously when they are taking it. This is a case where there's been a lot of concern."

Despite debate, controversy and skepticism, the Learning Survey remains an instrumental part of the Progress Report. This is largely because it offers a formal opportunity for the broader public to engage directly in holding schools accountable for elements related to conditions of schooling like teachers' expectations of students, effectiveness of school communication and levels of safety and respect. There were many complaints about the Bloomberg-Klein administration's lack of outreach to parents and communities as it reimagined the school system. The Learning Survey stands as one of the only ways that the DOE tried to communicate with teachers, parents, and students its new model for defining school quality beyond test scores.

According to a former executive at the DOE, the Progress Reports, "started from a basic desire to move away from any kind of input-based system where the thing that was being regulated from outside the school was pretty subjective." Using test scores alone to evaluate schools lead to "unevenness and the things that people paid attention to had to with things you could see in the newspaper or on a compliance report, but rarely was the focus on the kinds of learning needs through learning outcomes - indicators in terms of

the kinds of practices that you would want to see happening in a school to support a different kind of outcome." He clarified, "the central goal was to try and shift the focus onto that [different outcomes] by looking at outcomes in a much more robust way and also balancing that with inputs, but looking at inputs that were more linked in the research to strong instructional outcomes." This desire to shift the system's collective focus is reflected in the comprehensive nature of the Progress Reports.

In his book, *Lessons of Hope*, Joel Klein echoes much of this depiction. In describing how his team viewed the Progress Reports, he wrote, they should work "less like a club to impose consequences and more like a flashlight to illuminate what was working, identify where there were challenges, and figure out how to address them" (pg. 187). He continued, "What those nervous teachers couldn't know was that we were listening to them, and to our academy principals, who were telling us that test scores should be compared on a fair, apples-to-apples basis, and that they should be one element in our assessment of students, teachers, schools, and principals." Klein felt that these views "were echoed by many nonacademy principals as well. They also made a lot of sense to us." Given these views, the Progress Reports were designed to be complete, sophisticated and fair comparisons of 'apples-to-apples.'

Fair Progress Reports with peer comparisons

A second noteworthy feature of the Progress Reports was the methodology the DOE used for calculating schools' overall grades. Jim Liebman's team wanted the Progress Report to reflect a school's contribution to student achievement - "no matter where each child begins his or her journey to career and college readiness" (DOE, 2013). Aware of the role that demographics played in student achievement, the DOE emphasized progress with the Progress Report and calculated overall grades by comparing schools to peer schools. The DOE placed schools in cohorts among 40 other schools that served similar populations of students. Seventy-five percent of a school's overall grade was based on a comparison to peer schools (referred to as Peer Horizon scores) and 25% was based on comparison to all schools citywide (referred to as City Horizon scores).

The rationale for using a peer index to calculate grades was to "control for demographic characteristics of students so that the final score for each school has as little

correlation as possible with incoming student characteristics such as poverty, ethnicity, disabilities, and English learner status" (DOE, 2013). This was an attempt to make the system "demographically neutral" such that grades on the Progress Report "could not be correlated with zip codes," as explained by one former DOE executive. Another person who had been involved with the initial design, explained, "the rationale for going with a peer index for calculating grades was to make it fair - so you couldn't predict a school's grade based on student demographics."

The peer index, as described by one of the key architects, was used to communicate a set of principles using different tones. One tone could be characterized as harsh or menacing; the DOE wanted to send a message to principals that poor student outcomes could not be attributed to the students: "We were very clear, t's not a problem with the kids, this is a problem with what you're doing." The DOE used a more encouraging tone to "keep the principles of fairness and so forth." The peer index was meant to communicate these starkly suggestive messages.

One of the key designers explained that the decision to use the peer index was made, in part, to avoid some set of arbitrary accountability metrics. To accomplish this they decided to blend criterion and norm referenced models. He explained, " The big problem with criterion referenced is you set a line that you want everybody to meet and it's arbitrary and nobody knows whether you can meet it. That's the problem with criterion, and that's how the tests had been before, like Level 3, Level 4, they were bullshit, nobody knew, and it turned out that Level 3 was really so low it wasn't meaningful."

A strictly norm referenced model was also problematic because it "would essentially make zip code be determinative… But secondly, educators hate it because they say, well this is not an accountability system, you're telling me if I want to do better, get better kids, so I'm not going to pay any attention to it. Do something that helps me." The way the DOE managed this set of tradeoffs was to set the criterion-reference on how well similar schools and schools across the city did in the year prior; this way the bar was not arbitrary. He explained, "We weren't asking anybody to do something that couldn't be done, we wanted to set it so we were pushing you to do as well as the best had done but we weren't pushing you to do something that nobody really knew could be done."

Defining Quality Reviews in NYC

Beginning in 2006 and continuing under Mayor Bloomberg's control, the Quality Reviews in New York City involved a two-three day on-site visit of a school. Visits were conducted by a reviewer who typically wears another hat throughout the year as either a superintendent, network member, Quality Review Director at central office, or consultant in the system. The first task of the reviewer on the day of a visit was to meet with a school's principal to review the school's self-evaluation (SSEF) in which the principal outlined what he or she believed to be the school's strengths and weaknesses as measured by the Quality Review rubric. Using the same rubric, the reviewer would evaluate the school by observing classrooms; conversing with several constituent groups in the school, including teachers, students, parents, and community partners; and observing teacher teams engaged in collaborative inquiry. At the end of the visit, the reviewer conducted a feedback session with school leaders and network representatives in which they discussed the assessment. Schools are given an overall rating of either Well Developed, Proficient, Developing or Underdeveloped.

For each school reviewed, a reviewer writes a report, which undergoes rigorous quality assurance from the DOE to make sure that there is ample and clear evidence to support the reviewer's ratings. Once the report clears quality assurance, principals are given an opportunity to appeal. When finalized, the report is published on the New York City's Department of Education website.

A schedule for quality reviews is made available in the fall. Reviews were typically conducted between October and May. Annually, the DOE conducted approximately five hundred reviews. In 2012-2013, the DOE conducted 479 Quality Reviews (28% of the schools).

Site visits like the NYC's Quality Reviews were not completely new in the United States, but it was new that they would be used as accountability tools. Often referred to as site visits, the practice of qualitatively evaluating schools arose after the federal law, No Child Left Behind (NCLB) required States to create support teams to help schools identified as failing to make Annually Yearly Progress (AYP). Some States, like Ohio, responded by mandating site visits as a form of interventions for its struggling schools. In 2008, Ohio implemented their Ohio School Improvement Diagnostic under which

twenty-four officials conducted two-day visits to targeted schools (Jerald, 2012). Whether used as a targeted intervention for failing schools, or a part of school accreditation, schools typically treat the visits as bureaucratic hoops to jump through and the exercise usually has little influence over day-to-day practices in schools.

What is new to the American context is the systematic, routine use of site visits meant to identify *leading indicators* of school performance. Modeled loosely on public school inspections in Hong Kong and the United Kingdom, NYC's version of the site was an attempt to move away from bureaucratic hoop-jumping to promote the acquisition of continuous improvement practices and mindsets among those working in schools.

In New York City, the Quality Review was born with two important genetic traits that distinguish them as a different breed of site visits. The first is that Chancellor Klein mandated that *all* schools were to receive a Quality Review - regardless of history, past performance, or status. Legacy could no longer protect a school from scrutiny. All schools were to be held accountable for moving students. Secondly, a school's Quality Review rating was to be published on the school's progress report, which represented a symbolic broadening of the metrics used to gauge school quality.

Introduced in 2006, the Quality Review was intended to be a leading indicator to the Progress Report, which was seen as a lagging indicator of school quality. The Progress Report was based on students' test scores. Students take tests at the end of year, which makes the scores unavailable for mid-year corrections. But, the Quality Review was intended to evaluate the practices, systems, and structures in a school community and the extent to which they serve to strengthen teaching and learning (DOE, 2013). These were considered things that a school could attend to during the school year such that it could affect students' test scores. As one DOE executive explained, "if things are going well, we should see that in the outcomes later." The Quality Review was meant to name explicitly those "things" that need to be going well in order to see better test scores at the end of the year.

The Quality Review was never intended to be a prescribed manual or script. Rather, as one chief executive described, "I would call it a balance. I am not sure what they called it – a compliment to the quantitative stuff…I think the Progress Reports has been the main conversation piece. So you've got these big letter grades – like big scarlet

letters that schools get, but…it was very clear that the priorities, the values that were being prioritized by our system were about building a data culture in schools." The Quality Reviews were intended in helping schools develop the internal capabilities to respond to the "big scarlet letters." One of the architects echoed this point when he said, "I think part of what we were trying to do was to get that balance between the internal and the external. The external was there in my mind to drive and support and create some structure around the internal but it was not, by any means, to replace the internal or to deny that the internal existed; it was all in service of that…the question is, just how to get that balance right?" The Quality Review was the DOE's effort to answer that question.

The first Chief Accountability Officer in NYC compiled a team of people to create something qualitative. The team, explained one official, imagined the Quality Review to "most important thing you can do, but you need to be careful that it's not a bunch of educators going in and basically saying, well, this is how I would do it." Instead, the designers wanted a way to make explicit the process of developing professional expertise and strengthening pedagogy across diverse schools – where there are already norms and efforts to develop professional expertise. This was meant to encourage coordinated and coherent routines for collective problem solving. In addition, the tool, needed to be developmental and account for range and depth (or lack there of) of capabilities across schools.

An executive who managed the Quality Review explained, "Something like this [Quality Review Rubric] is useful for showing that it's not a prescribed way to focus on instruction, but to suggest the ways in which a schools need to be organized – the structures in place to look at and consider instructional quality…a high quality school does this – so, in that way, it was blending internal and external accountability – helping schools to be problem solvers."

Jim Liebman's team saw the Quality Review as a "crucial tool for creating more incentive on the part of principals to be strategic leaders," as explained by one of the team's members. Respecting principal's autonomy, the Quality Review, "didn't tell them how to do it, but it said, come up with a plan. It's your plan; you figure it out, but you gotta have a plan, and the plan should be tied to what you know about your students'

needs. You [principals] should be able to mobilize your program to meet that plan and then you ought to be able to figure out if it's working, and if it's not, do something pretty quickly to change."

Treating the Quality Review as a management tool allowed the designers to be "agnostic with respect to the particulars of instructional practice," explained a former DOE executive. Instead, the designers intended the Quality Review to emphasize a "particular view of how schools should be organized, and that view is sort of like continuous improvement - sort of process view," suggested a former DOE executive.

According to one of the architects, the team used, "every rubric I could get my hands on. They all had like 10 or 20 different things on them, but they all had being strategic on them." They looked at Reeve's, and the Principal's Association; the team traveled abroad to countries where they do school inspections to learn about the process. The point of departure for Jim Liebman's team was "stuff which is too compliance based. 'Are you teaching this curriculum?' and stuff like that." Rather, they wanted to measure "capacity not compliance," explained one executive.

In 2006, New York City's DOE contracted with Cambridge Education Consultants, a firm from England to manage the Quality Review process. Jim Liebman's team worked with Cambridge to customize the rubric. The first iteration was based on five Quality Statements:

(1) Gather data: School leaders and faculty consistently gather data and generate data and use it to understand what each student knows and is able to do and to monitor student progress over time.

(2) Plan and set goals: School leaders and faculty consistently use data to understand each student's next learning steps and to set suitably high goals for accelerating each students learning;

(3) Align instructional strategy to goals: The school aligns its academic work, strategic decisions, and resources and effectively engages students around its plans and goals for accelerating student learning;

(4) Align capacity building to goals: The development of leadership, teachers and other staff capacity is aligned to the school's collaboratively established goals for accelerating the learning of each student;

(5) Monitor and Revise: The school has structures for evaluating each student's progress throughout the year and for flexibly adapting plans and practices to meet its goals for accelerating learning.

Each quality statement had five to six sub-indicators, which were used to rate schools. The thinking behind this rubric was to "make the Quality Review mimic the thinking process that you'd want a good manager to go through because we wanted to train people, in a way, to be good managers. So, the first thing you do is figure out where you stand and then the next thing you do is plan. And then, you create strategies that suit."

The logic for the rubric was reminiscent of Total Quality Management, a management strategy popular among manufacturing and management executives in the 1980s. TQM suggests organizational learning occurs as members of an organization collectively commit to improving organizational performance through ongoing feedback loops and attention to system connectedness (Argyris & Schon, 1996; Fiol & Lyles, 1985; Senge, 1990; Levitt & March, 1988; Edmonson & Moingeon,1998; Ouchi, 2008). One of the central premises of TQM was that everyone in an organization - from the CEO to those working on assembly lines shared a vision for what the company was trying to achieve; every member of the organization was empowered and was seen as contributing to a service or product delivered to a customer. The connections between employee contributions and final product are easier to distinguish when the product is made on assembly line. The complexity of production in education is a little bit harder to untangle.

The team responsible for designing the Quality Reviews was well aware that TQM did not translate directly to the education context, but they believed that professional expertise would be generated from having teachers and principals engage in the procedures similar to TQM. One former executive explained, "Educators need to be able to get together in teams and be able to figure out what do we not know…and pulling all of that together and then allows you to sort of come up with a hypothesis about why instruction is failing for some kids, or if you want to put it another way, why this teacher is not able to master the imparting of this particular Common Core standard there?"

Another executive echoed this point explaining that the Quality Review was

meant to encourage educators to, "try to use things like root cause analysis and the five why's and anything that you can come up with to try to figure out it's not happening and then hypothesize a strategy that might work. Set some goals for yourself about what you think success would look like and then do it and see if works."   One member explicitly said, "if you go read about Toyota and its quality circles; this is exactly what their circles do." If assembly line workers, who do the same thing everyday can do this, "then you would think teachers can do it."

In their description of New York City's Quality Reviews, Stacey Childress and her colleagues (2010) also suggest that the Quality Review reflected a theory of action linked to organizational learning theory. Generally, the idea was to promote, "collective or macro-level" learning from feedback (student achievement data), which was supposed to lead to adaptive behavior based on that feedback. "Liebman sometimes described the rubric as a way to determine the degree to which schools were on the way to becoming 'autonomous problem-solving units' (Childress et al, 2010, pg. 13).

The working theory was that in looking at data, a school's problems would be evident, and then a team of teachers, encouraged by the explicit processes outlined in the Quality Review rubric, would come together and generate professional knowledge about instruction that they could then use to solve their unique problems. The tool was trying to account for local context – suggesting that schools would have unique problems, but in creating solutions by engaging in the data-driven problem-solving process outlined in the Quality Review rubric, the schools would develop sustainable problem-solving skills.

It was also imagined that engaging in the Quality Review process – the self-evaluation, the site visit, and feedback would provide schools with more material and data from which teams within schools could further enhance their learning to improve with data.

Research on schools' internal accountability systems had been around for a decade at the time that Liebman's team was working to design the Quality Review rubric (see Newmann, King and Rigdon, 1997) and the designers based the indicators on this research. Similarly, there was contemporary research being promoted like *Instructional Rounds* by Richard Elmore and Elizabeth City and other programs like DataWise out of Harvard that were promoting the use of structured protocols to foster school cultures that

support using data to drive instructional improvement. As such, the first iteration of the rubric was heavily focused on creating data-driven cultures in schools because it was imagined that this would lead to the development of stronger internal systems and professional expertise among staff.

The original architects used existing research and conducted some of their own in order to avoid incentivizing or encouraging superficial compliance-based responses in schools; instead they wanted to evaluate principals on how strategic they were in managing from improvement. For example, one of the original architects explained his process of trying to determine what to put in the rubric:

> …I went and read every single rubric that I could get my hands on. And they all had like 10 or 20 different things but they all had this in them, whether it's Reeves or the Principal's Association or something, they all had this thing about strategic planning, so in certain ways, I took the structure from what was already out there but the thing I did was to get rid of all the other stuff which is more to me like a compliance review and some of it was like, 'are you teaching the curriculum?'

While some of the research was about finding best-practices, other was meant to legitimate the tools. For example, one chief executive who managed the Quality Review explained that in the beginning they created a Knowledge Sharing Team and a Promising Practices Library to communicate the research that supported each indicator they selected. He explained, "It [Library] was sort of like the research behind the different aspects of the Quality Review, you can imagine just a big grid of all the indicators on the quality review, there were 20 at the time and then columns that were like, what's the research aligned to each indicator that really supports why this is important to schools, what does this look like in practice, sort of like, see it and then use it."

Cambridge used Liebman's data-driven, research-based rubric to evaluate more than 1,400 schools in NYC during the 2006-2007 school year. By the end of 2009, the consultants had conducted more than 2, 900 reviews. But, this was a costly venture for city.   A three-year contract with Cambridge was estimated to cost the DOE $19 million (Childress et al, 2010). In 2009, Chancellor Klein decided to make the Quality Review an in-house production. This lead to several changes to the process and rubric.

NYC's Version of the Quality Review

When Cambridge ran the Quality Review, every school received a Quality Review every year. However, the DOE did not have the capacity to bring the Reviews in-house and to conduct the reviews – especially because, with Mayor Bloomberg's small schools movement, approximately forty new schools opened each year. The DOE simply did not have enough people to be both reviewers and managers of the process for 1700 schools in the system. As a result, the DOE changed the schedule for when schools would receive Quality Reviews. The DOE reduced the charge for every school to receive an annual Quality Review to every school would receive a Quality Review at least once every three years; certain schools received reviews more often.

The DOE created selection criteria to determine which schools would receive a review in any given year. Schools that earned an A or B on the Progress Report combined with a rating of Proficient or above on the Quality Review would be scheduled to receive a review every third year - as long as the school maintained or improved their Progress Report grades. A school would be scheduled for a Quality Review in a given year if any of the following were true: (1) it received a rating of Underdeveloped the previous year, (2) it got a D or F or three C's in a row on its Progress Report, (3) it was in the bottom 10 percent of schools the prior year, (4) the school's last review was three years prior, or (5) it was chosen from a lottery of schools that had received a review two years earlier.

Changing the criteria minimized the number of quality reviews that the department needed to conduct in a year, which helped address their capacity limitations, but it did not serve the spirit of the Quality Reviews. The point of implementing the Quality Review, in the first place, was to promote the development of organizational learning capabilities in schools. Being restricted by capacity was not a good enough reason to stop implementing reviews. To address this gap, the administration created Alternative Quality reviews, which were to be managed by the networks. To ensure that this would be seen as formative, the results from Alternative Reviews would not to be published or used to make decisions about school closure, or principal bonuses; the Alternative Reviews were meant to provide additional opportunities to practice with the Quality Review rubric in a formal, but low-stakes way.

In 2010, the DOE introduced four types of Alternative Quality Reviews including:

a. New School Quality Reviews – (NSQR) is for schools in their first year of operation.

b. Peer Quality Reviews – for schools in their second year of assistance and for schools that demonstrate sustained student growth (3 A's on PR for 3 consecutive years).

c. Developing Quality Reviews - (DQR) for schools that received Developing (or below) on previous QR. This meant to help schools & Networks work together to strengthen the school.

d. Priority and Focus School Quality Reviews - Priority and Focus School Quality Review (PFQR) is designed to provide schools that have been identified by the New York State Education Department as Priority or Focus with the opportunity to work together with their network team to strengthen the school's instructional core, improve student outcomes, and meet the criteria for reaching proficient or well-developed on the Quality Review. In addition, the PFQR report includes a set of recommendations that will be used to develop the subsequent School Comprehensive Education Plan (SCEP).

The department continued to determine which schools would receive Alternative Reviews and disseminated that list to Networks and schools.

Under Cambridge, the rubric had twenty indicators - each with sub-indicators. As I described above, this version of the rubric was intended to promote the development of a data culture in schools such that the schools would become autonomous problem-solving units. "If you look back at the reports over the last five or six years, you see that one of the first two years, there were 37 indicators and 35 of them were about data. And it was clear that the whole point of what was going on was to create a data culture in schools; that was not lost on us," explained one DOE executive explained.

The reason why the DOE was pushing a data culture stemmed in part from its efforts to balance educator views on school improvement and the more business-based performance management crowd. In the beginning, the performance managers had more influence. The rationale that drove NYC's initial version of the Quality Review rubric was explained by one of the original design team members:

In the quality review, we're going to tell you a lot about how your school has managed it from a strategic management perspective and from that, we then want you to essentially do - in many ways we saw the quality review as kind of the Bible or the structure because what it said was, do you look at your data to see where the weaknesses are? If you do, do you come up with a plan to solve that? Thirdly, do you align all of your instructional activity to that plan? Fourth, do you align all of your development, professional development, youth development, leadership development to that plan? And fifth, do you monitor constantly to see whether the things you think you're doing are working and if they aren't change them? So that's what we wanted to be happening and we thought if we could get that happening, it would be good for schools. The quality review was both to measure whether it was happening but to provide a tool for schools to know where it wasn't happening and even in the particular places where it wasn't happening.

One of the consequences of pushing a data culture, in the beginning, was that far too many principals treated the Quality Review as a compliance task. One of the original reviewers, who served as a superintendent in 2006 and was trained by Cambridge to conduct reviews using the original rubric, explained that during that first year, principals would behave like the Quality Review was a curriculum audit. A team of reviewers would walk into a school and be escorted to a "conference room filled with binders for me to look at in a day and a half." The principals treated the Quality Review like a "dog and pony show" to get a grade. She continued, "I think it was a heavy shift for the system to move away from, 'we're not inspecting your curriculum; we're actually inspecting the systems and structures, the protocols, the rituals, the routines that you have here in the school to get a better understanding of where you're succeeding and where you're not. So that was big looking at it from the first year to maybe the third year. That was a big, big shift."

This point was echoed by another executive, "the rubric was overly focused on data and the use of data. In some ways, an unintended consequence, it wasn't intentional, there were behaviors starting to happen in schools to get ready for the Quality Review…it became just like test prep for the Quality Review. It was frustrating, and some cases, it

seemed like test prep for the Quality Review was getting in the way of what we were actually trying to do in the longer range school improvement work."

A second problem with the NYC's early version of the rubric was that it was too easy to get good marks. One executive explained, "when the rubric was first forming, the rubric was not so well defined and deep so you could do a quick prep and get a really good mark on that. We're talking about 2007, it was like you could clean up the house a little and tie everything under the rug and get a really good mark, maybe, but you could if you were strategic." A former network leader shared that he was able to, "to get folks ready…in my second year, having figured out the QR rubric, I got all thirteen of my schools a well-developed by kind of figuring out how to do this test prep work."

The Quality Review required principals to have a very different kind of attitude towards accountability, school improvement, central office, and their relationships with each. The Quality Review sought to use accountability as a school improvement tool, and to position central office, as the managers of the Quality Reviews, as partners with schools in school improvement. Principals' attitudes were a barrier to getting the Quality Review to work. For example, one executive explained, "I knew from my own work as an achievement coach in a network that, in some ways, the psychology was that it was a win, even though it wasn't a win for the kids, if the school's being untouched - even if the school's not doing what it needs to do."

A cluster team member explained, because, "We never really had a system like that here…principals didn't really understand the end goal of inquiry process; schools felt a lot of pressure to have a lot of data and create a lot goals and spent time creating these data binders and formatting and reformatting their goals, but it didn't more them to the next stage of inquiry…people tied that back to the Quality Review. They would say, 'it says here in the quality review that we need to do this. So why are you telling us its not productive?"

Similarly, a chief executive shared a story from the early days when one of the indicators measured the use of individualized student goals. He shared, "Word got around that reviewers were looking for, and that's the other thing that happens with this sort of - word whips around and for folks who aren't taking this seriously or don't think it's a real meaningful tool, they're like trying to find a quick fix, a test-prep way to get

through to get a good grade." This executive remembered that "kids memorized goals that weren't really meaningful to them, or had them taped to their desk to get through that part of the review."

These problems with the early rubric as it rolled out across the city started to influence the debates among the senior leaders. The initial version of the rubric drew heavily from the performance management side of the debate, but very quickly, it was clear that this version of that the management-driven rubric was quite limited in being able to trigger the changes it sought. The performance-management principles were insufficient "because schools didn't really buy into data-based goal setting," explained one chief executive in the DOE. Another architect echoed the point, "So many of the educators could not connect with this list of 37 indicators. It created so much cynicism about what the DOE was measuring as opposed to how kids are treated, or the quality of what's happening the classrooms."

A second unintended consequence of the early rubric was that there was a preponderance of Well Developed schools in the city. "The lay of the land was fraught with all sorts of problems and one of the biggest issues was that the vast majority of schools were getting high grades in quality review. You can check the data, but something like 85% of the schools got proficient or well-developed in the first couple years. That did not sit right with Joel Klein; it didn't sit right with Eric Nadelstern, who was gaining power, and it certainly didn't make sense to me," explained one of the executives. When I did check the data for 2009, 93% of the schools reviewed were rated either Well Developed or Proficient (top two ratings).

Seemingly more problematic was that among those schools that were rated proficient or above, "50% of the schools had 1.2, pedagogy of classrooms, was rated below proficient. And then, well-developed, it was something like a third below," explained a senior director. What this means is that schools that were getting high marks on their overall Quality Review rating actually had poor, inconsistent, or incoherent teaching happening classrooms. Indicator 1.2 measured whether schools "develop teacher pedagogy from a coherence set of beliefs about how students learn best" (NYC DOE Quality Review Rubric, 2015). To score "below-proficient" on this indicator reflected that teaching in the classrooms was not aligned, or coherent. The senior director

explained that it did not sit well with anyone, "that we're calling a school overall well-developed and proficient, but we're saying the classrooms aren't." As a result, the senior leaders set to change the Quality Review rubric to reward better the things that they valued. For example, one chief executive explained the changes were meant to reflect that the Quality Review was supposed to, "be not just an accountability tool, but like a values messaging tool." According to this thinking, the Quality Review needed to reflect that the quality of instruction across the classrooms mattered just as much, if not more, than the use of data.

With these implementation struggles, the debates about what to include in the Quality Review started to favor the educators. Educators on the leadership team were given more authority to change the rubric. This timing coincided with Jim Liebman, the first Chief Accountability Officer returning to his professorship at Columbia Law School, and Shael Polakow-Suransky, a former teacher and principal in the city, took over as Deputy Chief Schools Officer. The press release that announced the change described that Polakow-Suransky "began his career as a math and social studies teacher. In 2001, he became the founding principal of Bronx International High School, which has served as a model for the development of many of the City's new small schools. He has also served as a Leadership Academy facilitator, the Deputy Chief Executive Officer for the Office of New Schools, and the Chief Academic Officer for Empowerment Schools, which he helped build into the Department's largest School Support Organization."

With an educator-insider leading the ship, the newly created Division of Performance and Accountability overhauled the Quality Review rubric to focus specifically on the instructional core.

Drawing on the work of Richard Elmore and Ball and Cohen (1999), the DOE began to message the importance of attending to the instructional core in schools. The DOE defines instructional core as, "the relationship between student, teacher, and content" (DOE, 2013). Several senior leaders advocated for reorganizing the rubric, explained one director, "around Richard Elmore's instructional core theory. Basically, we wanted to say, we're not just trying to measure how well the school is organized to support students and them succeed. In fact, we know that there's inconsistency in the quality of what's happening in classrooms - a great teacher here and right next door, a

terrible teacher. So, we started aiming the rhetoric as this idea of the consistency and quality of classrooms across the school."

In September and October of 2010, members of the Academic Quality team (those responsible for managing the Quality Review) met with every network in the city and conducted several trainings with principals. The purpose of the meetings was to explain rubric's evolution and the new expectations that came with the changes.

With a power point presentation, the team framed the motive for changing the rubric (borrowing from a talk Richard Elmore gave earlier that summer) using research from Brian Rowan and his colleagues. The team explained that research suggests there is a greater variation in student achievement across classrooms than between schools, or among students (See Appendix D). This comparison set the stage for why they were going to start emphasizing instructional coherence over data culture in the rubric.

Much of conceptual framework for the new rubric was borrowed directly from Richard Elmore's workshops with NYC DOE staff that took place in several packed conferences rooms in Manhattan spanning several days in June that year. The DOE's presentation deck used during road shows, where central office staff explain changes to principals, networks, and clusters read, "Elmore agrees" next to the line, "variability of practice across classrooms within a school [sic] is higher than across schools."

The DOE was trying to shift principals' and teachers' attention towards the relationships between instructional leadership, school culture, teacher teams, and instructional coherence. However, they did not abandon data; rather, they connected data to instructional coherence. According to the DOE's presentation about the Quality Review changes, "effective data and resource use, professional development, goal setting and monitoring should be evident in the instructional core across classrooms. The accountability is in the task. If it's not in the instructional core, it's not there (question of impact)." This was an explicit move to root organizational learning in instructional coherence. The new rubric suggested that learning about practice required looking at data in practice - not spreadsheets. Evidence had to be visible in the classrooms, student work, and teachers' interactions. Instruction and coherence across classrooms mattered more. Few educators could argue with that reasoning.

Schools

In 2006 and 2007 when the DOE offered increased autonomy to schools, it was in exchange for stronger accountability. Equally important was the DOE's wish to be held accountable for system performance and improvement, which according to the core strategy involved marked improvements in equity in student outcomes, as well as, greater commitments to improving coherence and instructional quality. For these reasons and others, all schools in the city received an annual Progress Report that gave each school an overall A through F grade. Schools received annually a Progress Report that graded them along three dimensions: school environment, student performance and student progress. Schools started receiving Progress Reports in 2007, but the practice stopped in 2014, during Mayor deBlasio's first year in office; his administration changed the accountability tools significantly including eliminating the graded Progress Reports.

One goal for Mayor Bloomberg's graded Progress Reports was to offer principals, central office, and the public a way to monitor school performance and improvement. A second goal was to give autonomous principals, as the central managers of their schools, a management tool that measured their school's contributions to student outcomes "as accurately as possible given the different challenges schools face" (NYC DOE, 2012). Armed with the tool, the leaders were encouraged to verify and re-create metrics as this would permit them to understand better what they were being judged by; the logic was that then, this understanding would offer clear guidance for how each might improve his or her school's performance (NYC DOE, 2012). The comprehensive nature of the reports was meant to make monitoring easier, as well as, make it easier to target interventions for each school's specific needs. However, this assumed that principals possessed the capability to determine what were urgent areas to target from the mass of information now available. There was also an assumption that principals would possess the time, space, knowledge, and resources to decipher what required urgent responses from those that may have required more long-term vision, and the capability, to deduce the implications for future action and plans based on their interpretations.

For the autonomous principals, there was no mandate from some higher up that said they had to perform tasks associated with performance management; rather, it was implied that good managers, who led good schools would be able to perform these tasks -

to learn from and adapt based on data. There was an expectation that principals would be able to understand the problems made evident by the data and that they possessed the capabilities needed to envision and execute high-leverage solutions. The designs presumed that principals would vary in their skill with this, and as such, the DOE attempted to provide supports. Should a principal lack an aptitude for managing performance with quantitative metrics, she could turn to her colleagues with whom she worked, or she could turn to her network for support. As one architect explained, "That's why you had network teams; you had a data specialist who was meant to work with schools on their data."

The point of autonomy-accountability was for every principal to figure out how to solve problems evidenced by the results on the Progress Reports. Not knowing how to use the Progress Reports to your benefit was a problem the**,** principal, as a manager needed to figure out. What a principal could not do was ignore the Progress Reports.

Every October the Progress Reports were released to the public and all principals had to face their school communities – whether the school was above average earning an A, or whether it was a mediocre school with a C. If a school scored below a B, it was treated increasingly like a public relations issue for principals who would have to answer to parents and the community, but it also created cultural issues for principals who needed to find ways to rally staff who had been branded less than average by a less than desirable Progress Report grade. One network leader discussed how for principals, and the schools they led, the public nature of the grades became personal. She explained, "…what has happened in New York City because of the Report Cards, you kind of own it, it's a very public thing and you take it like your grade as a principal – cause even as we are all former principals (in the network), and we all felt the same way that if a school got a C, it was like 'omg, I got a C.'"

On principle, this was exactly what the designs were intended to accomplish: to have principals feel personally responsible for their school's performance. It seems like a fairly obvious thing – that a principal would feel responsible except that in education there are complex factors that influence student achievement, and in turn that complexity shields teachers, principals, and district leaders from taking responsibility for weak student performance. Under conventional models that based school quality solely on

student achievement scores, there were plenty of talented principals with strong moral compasses who led dismally performing schools without compunction because their schools, often located in struggling neighborhoods, faced a number of challenges that come with serving kids who live in conditions of poverty or violence, or have limited access to basic human services, or inconsistent housing arrangements. Similarly, there have been plenty of mediocre principals who led excellent schools that drew students from neighborhoods rich with social and academic capital. Conventional models were limited in their ability to distinguish what students contributed to student outcomes, and what leaders did. In some cases too much credit was placed on poor performing students, and not enough on high performing students. The conventional models allowed for muddy waters, where it was difficult to see clearly the linkages between what a school provided and how students performed on tests.

The Progress Reports, by contrast, sought to push people to look more closely at schools' contributions to student achievement and progress, but they attempted to do so fairly. According to the DOE, the Progress Reports were meant to "produce outcomes that are minimally correlated with socioeconomic status, Special Education populations, or other demographic variables" (NYC DOE, 2012). To achieve this, the DOE created a formula for comparing schools with similar populations and weighted more than three-quarters of a school's grade on a comparison to its peers. This created a reversal of sorts – especially for the top and bottom schools.

For persistently low performing schools in NYC, the Progress Reports created a way for a school to be judged as to whether it was doing everything within its control to help students achieve. Student progress on tests counted for nearly sixty points of the one hundred points that comprised a school's overall grade, where as student performance accounted for twenty-five points. Progress was twice as important as students' absolute scores, which meant the Reports could "reward schools for moving kids that may not be proficient, but are doing the work to get kids on their way," according to one of the original architects of the Progress Reports. He added, "because no one state or federal was doing that. So, the original design was saying we want to make sure we are raising the performance of all students."

Persistently low performing schools were encouraged help students make

progress even though the students may come to school two, three, and sometimes, six grade levels behind where they should be. Conventional systems did not work this way; they counted absolute scores more than growth, which made it very difficult for schools serving larger populations of disadvantaged students to get recognized for herculean efforts. The Progress Reports incentivized schools to care more about progress so those schools that had been plagued by decades of poor academic performance, or a culture of low expectations needed to reverse expectations and focus specifically on the lowest performers (Hemphill, 2010).

For the historically top performing schools, the Progress Reports did not necessarily reverse expectations; rather, they introduced skepticism into longstanding ways of thinking about what constituted a quality school. This definitely affected schools disproportionately – especially those that in the past may have felt very secure in their standings. One senior executive at central office explained that for many schools the shift resulted in different kinds of outcomes for previously top performers, which was seemed counterintuitive, especially for parents. He explained, "historically high performing school don't always do well [on the Progress Reports], like if in New York City, the way you got to be a good school depended a lot on who entered your school and so lots and lots of schools created some kind of screen. Sometimes the screen was the neighborhood zone, sometimes the screen was a gifted program, sometimes the screen was an arts sort of audition or something like that, and still I think in most people's minds even outside of New York City when they think of a prestigious school it's a school that's hard to get into and the kids are smart who go there and there's this test or some kind of admissions process to get in. And that's true at the university level too."

This kind of thinking was the prevailing paradigm under conventional models, where to many, the way to look good on accountability metrics was to attract better students and "that doesn't actually correlate much at all with whether the school is good cause a school being good means it's moving the students who are going there," offered one DOE executive. Old input-based models, privileged those schools that could draw 'better' students, whereas schools located in high poverty areas, where educational outcomes, as measured by standardized tests, did not come so easily – these schools were hit doubly hard by conventional models.

The spirit of the Progress Reports was to create more equity in the system, and as such, the Progress Reports "started from a basic desire to move away from any kind of input-based system where the thing being regulated from outside of the school was pretty subjective and dependent on the supervisor working with that school and whether that was a district superintendent or regional superintendent…there was a lot of unevenness" and those differences could influence who was in your school's population, as explained by a chief executive in the DOE. He further explained, that NYC DOE was seeking a more even measuring stick that would help parents, educators, and the DOE recognize distinctions among institutions and how each was moving its students. He suggested that before the Progress Reports, under the old school evaluation system, it was possible that a good school looked good even though, "it just sort of keeps them [students] at exactly the same sort of level they came in at and it's kind of interchangeable which institution they went to because they would have been fine no matter where they went…impact isn't visible in those outcomes." The point of the Progress Reports was to distinguish school's impact and better understand what schools were doing because, as he explained, "well there are some schools that take the same kids both at the high performing and low performing end that have different outcomes. Some kids do a lot better from the same background and even when there are concentrations of high poverty or concentrations of academic strength, different things are happening in those schools."

The move to emphasize school's impact meant that there was the burden of proof placed on the schools. This burden existed for those schools that in the past did not really have to attend to accountability metrics because, in many ways, those schools rarely faced any threat of sanctions, which was, partly, due to the aggregate talent of their students. With the Progress Reports, these schools were in the position of having to prove that they were good and better than their peers at moving their already high performing students and that they possessed school climates where students could thrive and students' parents and the staff were satisfied with the school. While research has suggested that both top and bottom performing schools pay attention to accountability metrics (i.e,. Spillane & Diamond, 2004), the Progress Reports incentivized schools to care for and attend to things that in the past may not have been high priorities. The Progress Reports set a new expectation for what schools were responsible for – and this

was true for all schools.

One senior executive in the DOE explained that this shift was "critical." He explained:

> I think that for accountability to make the shift from simply focusing on student performance to focusing on student progress was a critical piece because the idea was that – for the kids that come to your school – how do you continually improve and develop them? I think the progress report in capturing that I think surfaced what are some skills that schools previously had been perceived as lower performing schools because their kids came in so low, but they actually did amazing work with them that was never really captured. And then the other piece was to start to disaggregate data and to say you know what let's look at your student body and who are your kids? Because there are kids who have greater challenges to be successful than others and to capture that and recognize that and acknowledge that in the work of schools who do great work with particular groups of kids to be successful is an important piece. And it also brought to the surface that schools need to attend to all of their kids.

This shift was not immediate, but did occur as the Progress Reports continued to clarify expectations for schools and parents, principals and central staff started to discuss school quality routinely in these terms.

Another way that the DOE incentivized all principals to pay close attention to the Progress Reports was that they decided to award principals performance bonuses, if their schools met or exceeded performance targets on the Progress Reports, which was part of the administrator's contract. In 2012, 275 schools qualified to receive performance bonuses and seventeen principals received the top award - a $25,0000 bonus (Phillips, 2012). Similarly, the Progress Reports, in combination with the Quality Review were used to remove principals whose schools persistently performed below targets. This practice was not without controversy or criticism especially given the volatility of some of the Progress Report scores in the middle and elementary schools.

The Progress Reports for the high schools were based on more measures (i.e., graduation, college-ready indicators, 4 Regents, etc), but the middle and elementary schools were based on state test scores. Hemphill (2010) reported that The Center for

New York City Affairs in an evaluation of the Progress Reports found that "schools may go from the very bottom of the city's rankings to the very top – and vice versa – in just one year. The Center found that more than half the city's elementary schools and middle schools had swings totaling more than 50 percentage points in their rankings over three years" (pg. 2). This was something that the DOE was attending to and made changes to the Progress Reports in 2010 in an effort to reflect more accurately what the school was contributing to the score, which would made decisions based on such data, ideally fairer.

As one former executive noted, "Every year, they take a hard look at their instruments." These hard looks were meant to make sure the tools were accomplishing the task of holding principals accountable for school performance in fair and accurate ways. For example, the DOE's guide for the Progress Reports explained, "These changes have strengthened the measures, enabling them to better capture schools' contributions to student learning. The DOE has also made changes to mitigate unintended consequences and to align with shifting local and national priorities, particularly a focus on college and career readiness" (DOE, 2013). The DOE communicated changes they made to the Progress Reports, why they made the changes they did, and throughout, refused to lower expectations for school leaders.

The DOE created a methodology to accurately measure a school's contribution to student learning that relied on very sophisticated statistical calculations; creating the methodology was an enormously complicated task. The expectation was not that all principals would understand how to construct statistical models; rather, the expectation was that principals would have enough working knowledge of the methodology and skill with interpreting the Progress Reports that they would be able to manage improvement efforts and anticipate what decisions and policies would provide the greatest leverage to improve student progress within their unique school communities. To accomplish this, transparency was paramount.

The DOE created annual Educator Guides and distributed these to schools and the public. The Educator Guides offered an in-depth, user-friendly overview of the Progress Reports including the purpose of the Reports in the system, their relation to the Quality Reviews, and a breakdown of all of the components. The Guides also offered explanations of the methodology used to calculate each of the scores and communicated

what changed from year to year. These Guides were housed on the DOE's website and easily accessible. In addition, the DOE's Office of Performance and Accountability, the office that managed the Progress Reports conducted Road Shows where they met with networks and principals to communicate the changes made to the Progress Reports each year.

There was an incentive for schools to pay attention to the Progress Reports and Quality Reviews - in part because there were steep consequences attached to schools' ratings on the Quality Review and Progress Reports and principals' tenure was connected to the outcomes. Getting lower than a Proficient on the Quality Review meant that your school would be subjected to go through a Developing Quality Review the following year. If your school earned less than a B for three years a row, your school would be subjected to possible interventions including replacement of school leadership, TAP plans, or faced possible closure.   Given the bi-level accountability scheme where as networks, too, were held accountable for their schools' results, the networks also helped schools pay attention to the Progress Reports and Quality Reviews.

Unlike prior administrations, Mayor Bloomberg and Chancellor Klein used school closure as a lever for school improvement. These policies were extremely controversial, but the leaders did not shy away from controversy, if it seemed to promise better outcomes for students. Between 2002 and 2008, the DOE closed twenty-three large failing high schools and replaced them with 216 new small high schools. In a study published in 2012, Bloom and Unterman reported that these changes led to improved and sustained 4-year graduation rates. Bolstered by positive effects for students, the practice of closing failing schools and opening new ones continued through Mayor Bloomberg's time in office.

During a conversation, a former chief executive in the DOE, reflected on the practices of closing schools: "Today, in my home borough of the Bronx where I couldn't find a middle school or high school to send my daughter to and wound up sending her to public school in Westchester…Today in the Bronx there are over a hundred high school choices alone available to parents and children, most of which are doing exceedingly well and many graduate between 70-80% of their students. In fact, during that period from 2003-2011, in eight short years, what had been a 50% graduation rate firmly in place for more

than 50 years increased by 30% to 65% and I think the leading reason for that increase was the closure of well over a hundred large failed schools and the fact that we opened well over 500 new small schools." Another executive explained, "…school closure has been so powerful because it's the only mechanism allowed within the teachers' contract to replace staff. There's no other mechanism available when there's a group of educators in a school that have historically not succeeded." Consequences for schools for repeated poor performance were immediate. As such, people working in schools and the network teams that were charged with supporting the schools paid close attention to what the tools were measuring and how the criteria changed from year to year.

In summary, the accountability tools changed what people in the system paid attention to. This point was expressed by one chief executive in the DOE when he described his opinion of the accountability system. He shared:

> I think the approach to accountability is thoughtful, controversial, and still evolving. The thoughtfulness comes from a place of – even as a principal of 13 years, the accountability on me to do my very best came principally from my own self-driven stance. The accountability exacted by the system was nonthreatening so for so long schools could languish and not do well and there would be no consequence unless the state decided it would close it as a result of you being a SURR school. Then the accountability under this administration [Bloomberg] came into play where through the accountability two things happened: accountability force people to pay attention to aspects or areas of concern like areas with students with special needs and English-language learners – the bottom third in your school. Many schools didn't know what the bottom third meant. Even at Stuyvesant, you had a bottom third. The bottom third may be achieving much more than other schools, but they're still the bottom third of that school. So it brought into focus things for people to both understand and focus on in terms of moving the student achievement, and it also forced leaders to take a look at themselves because it moved from you just managing a school to you being and becoming an instructional leader with the ability to stimulate and support others so that you become a team that can really move the teaching and learning in the building. So all that happened through the accountability.

Regardless of whether a principal led a school with high performing students, low

performing students, or a mix, she was called on to move students by moving teaching and learning; the combination of the tools was used to focus attention on both aspects – moving students and moving teaching through stronger management of the conditions. It was an intervention that did not tell you what to teach your students, or how; rather, it put the burden on schools to be excellent at the process of determining what to teach to whom and why.

The DOE created the two tools because, as one of the original architects of the Quality Review explained, "the new leadership wants us to think about using accountability to create pressure about being a learning organization as a school. So what would that look like? How do you create accountability supports or at least policies and tools that help people feel like they're being supported to move kids and not feel like they are being chased or hounded?"

The challenge of balancing learning with pressure was constant for the designers; they were conscious of the challenges that they were trying to solve, and aware that the designs may be creating new problems for schools. For example, one of the original architects commented, "I think having a technical foundation is really important, but if those teachers are very sensitive to how successful they're being, they'll quickly find that their technical knowledge is limited, there are lots of things that they don't have the knowledge on and even in the subjects where they do think they have it, it turns out they don't. And I think the common core, cause it's creating these more demanding standards, is making it clear that the technical knowledge that's there is, even for good teachers and good schools, less than people know, and it's a big issue."

Yet, the accountability tools were limited in helping people acquire this necessary knowledge. The DOE relied on the networks and Citywide Instructional Expectations (CIE) to help address that "big issue," but it was one that was brought to the surface by the combination of the tools that were meant to spotlight a school's strengths, as well as, its weaknesses.

The combination of the tools was meant to encourage an inquiry approach towards improvement and articulated the developmental phases and indicators of progress in any particular school. However, getting a school to change its culture and practices is extremely difficult – especially for low performing schools, where a

preponderance of inexperienced or substandard teachers are ill equipped with necessary technical expertise to fulfill the charge of teaching students who may face tremendous challenges in their lives outside of schools. Leaders in these places must contend with intensified pressure, while they, often, have fewer social and political resources to be able to build upon whatever assets may exist in the school.

The designers sought a way to balance the "performance management aspects of the whole sort of reform and the Progress Report" with "creating internal accountability," as explained one of the original architects. He explained further that the DOE was trying to find a way to "sort of rely on the professionals to be a little bit more in charge of holding themselves accountable and holding each other accountable and things like that." The point was to require schools to use these tools, and an inquiry approach, to force interactions among the professionals to solve problems related to weak instruction, but this proved to need time for people to get used to these kinds of ideas – especially within a context where things may not have been fair, or focused on formative learning in the past.

For example, one executive explained, "I think it does feel like we have put too much pressure and over-simplification on principals around these letter grades – so, that we sort of distracted too many people from the learning part." He described a principal who was able to get past the distraction, but the executive suggested that this principal was an exemplar. The executive shared, "And this guy figured out how to kind of, he felt the pressure, but instead of doing something that's potentially not useful for kids and his teachers, he's figured out how to use the data and the Quality Review lens to move the agenda that he wanted to move and focus it. And I don't think that's a common story, unfortunately." Clearly, another study is needed to tell the story of the implementation puzzle for the Progress Reports and Quality Reviews.

Central office

Mayor Bloomberg and Chancellor Klein reinvented NYC's central office (see Honig, Copland, et al., 2010; Hill, 2010; Fullan, 2014; Klein, 2014; Nadelstern, 2010). The formal hierarchy changed several times during Mayor Bloomberg's tenure, although, it did not expand; rather, it shrunk considerably. Nearing the end of the Mayor's time in

office, "the administrative bureaucracy outside of the schools is a fraction of the former size" (Hemphill et al, 2013, pg. 5). A review conducted by Parthenon reported similarly that the Mayor reduced the layers of bureaucracy in order to devolve financial resources to schools. By consolidating support services, the city saved $85 million, or 32% of operating costs (Parthenon, 2013).

One of the changes was the type of work conducted by central office. The Division of School Support was created to accommodate many of Mayor's initiatives, and within that was the Office of School Accountability. The Office of School Accountability would become the Division of Accountability and Achievement Resources – showing the DOE's commitment to balance accountability with support. Eventually, the division would be renamed to Performance & Accountability.

The DOE had to develop the capabilities to manage all aspects of the Progress Reports and Quality Reviews. In the early days, there was a significant push to build a team of people capable of leading the designs, as well as, a group to execute the plans on a day-to-day basis – at a scale unmatched in the U.S. From those I spoke with, there were three important aspects that allowed for the Office of Accountability and Performance Management to develop as they did.

The first is that the scale and magnitude of the changes they sought presented challenges, but it also created opportunities. As one of the original architects expressed, "And it [scale] could work both ways." He explained,

> …the team of people that I had working for me and that were engaged in all of
> this at central were exceptional and wanted to be in New York and wanted to be
> in this reform and wanted to be around what the mayor and Joel Klein were doing,
> and it's harder to attract as deeply concentrated a bench or pool of people as we
> had. Part of that is not just that it's New York and sexy and all that, it's partly as
> Joel, first thing Joel ever told me when I went to work for him is he said, 'Judge,
> your success by the quality of people you are able to attract to do this with you,'
> and I think that was true. We were attracting people because good things were
> happening and smart things were happening and people were really committed
> and thoughtful about it. But that's a human capital infrastructure that other places
> might not be able to do. I've tried to help folks in Connecticut, for example, bring

people there and it's much, much harder, just geographically to have people go there. Albany where it's just different, Detroit. And I guess I would say that I guess maybe it's the reverse point or the reverse side of the scale-point that I made, there are many, but at the time there were some really good principals and teachers. I'd get together with some of these principals and I'd think, these folks are just as interesting and smart and creative as my faculty at this fancy university. So I think that was a real resource for the system.

The DOE was able to attract a 'deeply concentrated bench' of talent to help lead and manage the bold reform.

When the Chancellor and Mayor took over, the talent that they hired, in some ways was controversial because many were not from education, or people who had been in the system. According to one supervisor, "well, in the beginning, the policymakers here at Tweed – the ones that Joel Klein…hired were attorneys by trade. Smart set of folks, but I look at what the charge was for Joel – Joel Klein wanted to come in and raze anything and everything that would prevent him from promulgating the new way – the fresh way. So, you want to bring in people that are less inclined to say no."

To rebuild central office and recreate how the district functioned, they relied on people who came from business and law sectors, or hired those who had been in the system, but were the people working on the fringes – people like Eric Nadelstern who was a champion of autonomy and the Empowerment Zone, but was seen as someone who "broke the system" during his time as Deputy Chancellor of School Support and Instruction. Eric broke it and then Shael Polakow-Suransky, a successful principal and former teacher who had taught at one of Ted Sizer's schools, stepped into the role as Deputy Chancellor and had to rebuild central's infrastructure and frameworks to support the implementation of the Chancellor's system-level reforms.

Among the new executives working at central office there was a fierce commitment to changing accountability in the system – shifting from transactional bureaucratic roles, to a problem-solving approach focused on expanding high quality schooling in the city. This created a new culture in central office that focused more on the adults in the system being responsible for working as a team to get things done, which became a resource. For example, a former director explained, "I guess, I'd like to say

that was the culture and, I guess, the source of that culture was definitely – it was not acceptable to have a problem flagged without a resolution in a reasonable amount of time and I mean I enforce that with the people who report to me. Shael enforced that with me and I think the Chancellor enforced that with Shael so that was…New York City." Another former executive compared it to a dynamic, fast-paced television show that depicted working in the White House: "Everyone coming in was new and brilliant and hard-working, and once that was in place, and it was a given, we could pull it off – it felt a little like West Wing – a little like that. The scene was chaotic and they got shit done."

Under the Mayor's and Chancellor's leadership, central office became, in the words of one former DOE executive, "mission-driven start-up." He explained, "The DOE at the time, and I think still is, is it's a really dynamic environment, which is a strange thing to say about the largest bureaucracy in New York City, but at least in the kind of management levels of the department, it really has the feel of a mission-driven start-up. And so what that means is there is a lot of collaboration within units but also across units, there is a lot of enthusiasm and people are really eager to share what they know and kind of work with people that so that it advances the overall project."

The resources that the DOE needed to make this work included new people with different dispositions and skill sets – many who were new to education. Central office also had to acquire technology and new processes for information management, training, production of reports, quality assurance, evaluation of the tools, and perhaps most importantly, data management. In the beginning, the DOE turned to several different consulting organizations. The DOE had contracts with McKinsey Consulting, McGraw-Hill (Grow Network), Cambridge Associates, and others. Being able to rely on consultants to help incubate programs and tools was an important part of central's transformation. One former executive explained, "There was never – there was never a question, 'do we have the enough money to pull this off?' In terms of resources, we had whatever we needed. If we needed designers to come in and build us a tool, we got it – and a company from England to develop the Quality Review for 4 – 5 – 6 billion? Then you got it. There were no resource constraints at all."

Over time some things, like the Progress Report production that was managed by McGraw-Hill initially, and the Quality Review process, which was managed by

Cambridge, did get brought in-house, while other things like the Learning Survey remained outsourced. The DOE kept its contract with KPMG Consulting to manage the Learning Survey because doing so allowed parents, teachers, and students, who took the survey, to feel their responses were confidential and for the survey to be a legitimate instrument. Another reason the DOE decided to outsource the Learning Survey, as explained by one former executive was because, "it's the largest survey in the country other than the U.S. Census and it costs one million – 1.4 million a year, which is like nothing."

Building the ARIS data management system, which tracked every student's grades, state test scores, credit accumulation, and other demographic data was a huge undertaking. Central needed to hire data specialists, equipped to create a massive data infrastructure and write code. According to one former executive director, what worked in NYC's favor was that it had "good student IDs so you could do all of the merging of files" and track students longitudinally. Central office had to develop the capabilities and hire talent that could build the necessary systems, as well as, develop systems to turnkey and communicate with people working in the schools who were responsible for inputting data.

Central office was committed to helping schools develop continuous improvement practices, but it, too, was developing systems and practices in order to improve continually the Progress Reports, Quality Reviews, and related initiatives – including a network structure, Common Core implementation, new teacher and principal evaluation systems, Citywide Instructional Expectations. Central had to stay on top of the design and implementation of each of these, but it also sought to align these in order to build coherence and stability in the system. To track all of this, central created quality assurance and communication systems in the service its own continuous improvement practices. For example, central created Road Shows in which members from the Office of Accountability would meet with principals and networks "to see what they were learning from it [Progress Reports]," according to one executive. One former executive shared that the Office of Accountability would conduct seventy sessions twice a year with people in schools and networks to learn from how people in schools and school support were using the tools. He shared, "this gave a human aspect to the work – so it

didn't always feel like crunching numbers, and a good way to engage with primary stakeholders." While relationship building was a critical aspect of implementation, it was not necessarily addressed by the designs for the tools.

Central office managed its own continuous improvement practices, and managed pilots for many of its initiatives (another perk the city's size; scale permitted them to roll things out on a smaller, but significantly heterogeneous group of schools), which allowed them to learn from the pilot and make adjustments before rolling something out city-wide.

Throughout, central office tried to communicate its strategy and what it was learning to those internal to the DOE and those external, including the public and educators. However, central also had to deal with the controversy surrounding its strategies, and "noise" from detractors who disagreed, or were dissatisfied with the disruption caused by the Mayor's and Chancellor's plans for transforming the system, as explained by one of the original architects.

Finally, the reforms required, in many ways, for stakeholders in the system to unlearn, or revise their ideas about what constituted a quality school. The strategy was to give parents better information about schools, but that information, at times, created a tension with their perceptions. Central office had to manage those public relation issues. The Progress Report presented a particular challenge on this front because while they were more accessible to parents (and occurred every year), they were also slightly confusing to parents and the public. The grades on the Progress Reports were calculated by comparing peer schools, which meant that sometimes an A school may have students who are performing well below grade level, if all of other schools in the peer group are doing even worse. Similarly, a school where a majority of the students are scoring well above grade level may get a B or C on the Progress Report because the school is not doing as well in climate or progress as its peers. For example, one executive explained,

> So when you have a grade that's based on progress, you can have a school that's getting an A because it's taking students from being very very very far behind grade level to being very far behind. And, according to our system, because we measure progress, you can still get an A when you made a lot of progress even though you're far behind where you need to be…that is consistent with your goal

to measure and reward growth, but it is also very confusing to parents because when they think of an A school, they think of a school where kids are meeting the standards in each grade level in the school. And the opposite also happens, too, schools that are given low grades that aren't making much progress, but have high proficiency kids.

This created a tension that central office had to manage – namely, that the accountability tools were meant for several audiences – some internal and some external that needed different information. There were tradeoffs in making one set of tools for all audiences to use.

This point was echoed by another executive, "that doesn't make a lot of sense to the public because they think a good school is where students are doing well, and that's not crazy – there's something to that, but in New York City, the decision was to emphasize the performance-management aspects of it and so the result was a fair amount of confusion in the public and what the evaluations means for them." Central office was responsible for trying to help the public and parents clarify and sort out instances where they may have been confused so that the tools would, indeed, be useful to them.

Environment

"The educational reforms in New York City's public schools under the administration of Mayor Michael Bloomberg and Chancellor Joel Klein from 2003-2010 were among the most ambitious of any large scale urban system in the country…The sheer scale of the city school system, with its 1.1 million students and over fifteen hundred schools, and the ever-present drama of New York City politics might be enough reason to engender interest in any period, but under Bloomberg and Klein it was the specific collection of reform policies and actions that captured national attention. These included a combination of mayoral control and fundamental restructuring of the system, the introduction of accountability practices, test-based metrics, and human capital models from the business sector; and the implementation of a panoply of reform strategies currently in vogue, such as small secondary schools, public school choice, and the closure of so-called high school dropout factories. Indeed, New York City seems to have drawn together many of the threads of what is emerging as a national education agenda." – O'Day, Bitter, Talbert (2010)

When Mayor Bloomberg took over control New York City's schools in 2002, the country was just beginning to make sense of the federal regulations in No Child Left Behind (NCLB) in 2002. There was an ongoing urgency to use student achievement data to

inform instructional reforms, but it was not clear what reform levers were best to pull. Compelled to make education reform and fixing the city's schools clear priorities, Mayor Bloomberg decided to transform the entire system; instead of going for incremental change with one or two levers, he pulled them all.

While revolutionary for its scale and reach, the Mayor was not alone in this thinking. In 2002, several cities and districts like San Diego, Chicago, and Duval County sought to reform their systems, as well (see Talbert & McLaughlin, 2003). Taking the system as the unit of change, these districts sought to develop the districts' capabilities and structures that would support improvement in teaching at the school and classroom level. Past reforms often focused on the efficiency of a district, but this new wave was seeking to promote system learning in the service of improving teaching and learning in schools (Honig et al, 2010).

One of the key challenges the system reformers faced was that neither practice, nor research, had determined which were the essential elements of systemic reform; there were few lessons or models for Mayor Bloomberg and the original architects in NYC that could help them convert complicated ideas into action on the ground that was relevant and meaningful for those who would have to run the system long after their terms were up. Mayor Bloomberg and Chancellor Klein had to figure out how to realize system-wide change in even, consistent, and sustainable ways – and they had to do this at a meteoric pace given the Mayor's limited time in office (no one could have anticipated that he would have eleven years in office; they had to work in four year increments).

Mayor Bloomberg and Chancellor Klein sought to manage complexity by putting forth a performance-management based set of reforms that relied on accountability for results paired with innovative supports, which created challenges, controversy and capability problems. Layered on top of these challenges, the plans also had to account for shifting priorities from New York State and national agendas. Throughout his tenure, Mayor Bloomberg had to work with the State dealing with a range of things from charter school caps, laws for teacher and principal tenure, and school funding.

One of the central topics that the State and city had to cooperate on was how to identify and deal with 'struggling schools.' NCLB in 2002, Race to the Top in 2008, and NCLB Waivers in 2013 required New York State to specify plans for how it was dealing

with its lowest performing schools – some of which were located in New York City (particularly District 7 in the South Bronx). Yet, the city and State used different metrics to measure school performance, and had a different set of interventions for persistently low performing schools. Central office had to figure out how to reconcile the differences, as well as, manage the impact discrepancies between city and State accountability systems had on schools – especially its lowest performing schools.

Managing this piece reflected several changes to the city's accountability policies including changing the criteria for Quality Reviews. For example, New York State Department of Education used site visits as an intervention for schools targeted as persistently low performing. During the early years of the Quality Review, the city did not attend to this State's site visits, and as a result, low performing schools were getting double visits, which created an unproductive burden on the schools. The city changed its criteria and tried to align with the State's schedule so that schools would only have one visit – performed simultaneously by the city and State. In doing this, the city also tried to communicate to schools how the city's indicators aligned with the State's, or at least where there was congruency so schools did not feel as though they were being evaluated on different measures, which would have complicated matters and caused confusion as schools tried to respond to feedback from the State and city.

While this section began by focusing on the time when Mayor Bloomberg took over control of the schools, it was not to suggest that this was the origin for all of the reforms. Rather, the reality, as was alluded to in other sections, was that many of the steps taken during the Children First Reforms were connected to educators, organizations, and ideas that predated the Bloomberg-Klein administration (Gyurko and Henig, 2010; McDonald et al, 2013). The environment contained several seedlings and pockets of proponents who were seeking, and had been working towards similar kinds of changes. For example, one of Mayor Bloomberg's key initiatives was the small schools movement with the help of significant funding from the Bill and Melinda Gates Foundation. With the small schools movement, the portfolio management model would likely not have worked. Joe McDonald (2014) highlights how the small schools movement had firm roots in NYC dating back to small schools that opened in response to critiques like Herbert Kohl and Jonathan Kozol, who portrayed the injustices of schools

that served poor and minority students. There were several reform efforts like Deborah Meier's Center for Collaborative Education (CCE) that was affiliated with Ted Sizer's Coalition of Essential Schools (McDonald, 2014). Meier's efforts to open multiple small schools and break up large school and move students to new buildings to avoid "carry over of old culture" would foreshadow what Bloomberg would create with the portfolio office that served to position new schools, including charters, where they were most needed (McDonald, 2014, pg. 120). However, the environment also had several well-funded and established detractors who had benefitted a great deal from the old interest-group politics that ruled before Bloomberg dismantled and disrupted connections (Hemphill, 2010; Klein, 2014; Nadelstern, 2013). These groups put up a significant battle throughout and became even more influential when Mayor deBlasio took office in January 2014.

Conclusion

In this chapter, I began by highlighting that states and districts are mandated to publish scorecards about school performance in order to help parents make decisions about where to send their children and to help the public have a better sense of the state of schools. The policies mandating the reports treated the scorecards as a means of informing and motivating changes in school performance. Yet, most scorecards rely mostly on lagging information about schools – student achievement scores, and graduation rates broken out by student sub-groups. In New York City in 2007, Mayor Michael Bloomberg and Chancellor Klein implemented a radically different set of accountability tools that were based on what they, and others architects referred to as lagging and leading indicators of school performance (Childress et al, 2010). The rationale for creating these tools was to fit within a broader movement towards performance-management at the system and school levels.

NYC's tools, in combination with the other structural changes like granting all principals autonomy and providing external support partners, were meant to shift fundamentally the culture of the system. Joel Klein, the chancellor, was reported as saying, "I think if you don't change the culture of public education, you're not going to change the outcomes materially. A culture that doesn't focus on performance is a culture that does not work" (in Childress et al, 2010, pg. 90). Drawing on organizational

learning principles, the Progress Reports and Quality Reviews were intended to be critical resources in shifting the culture across the system, and in all schools, to focus more on performance and managing for improvement than what had been done in the past.

This analysis has several implications for practice and research on alternative accountability models. First, the decisions that the original architects had to make, and the debates about how to shift culture system-wide towards empowered schools capable of managing for improvement, while also working with the vastly diverse set of schools and the capability challenges created by the shift they desired to make meant that the architects had to balance top-down changes with bottom-up, on-the-ground changes in schools. Liebman and Sabel (2003) noted, for systemic reform to move from an idea to an imminent reality, it had to balance the general rules and frameworks for instruction with changes in professional knowledge among teachers and structural changes in the organization of schools and districts.

These changes had to be orchestrated and coordinated to account for the complex nature of the changes that they sought. If the tools were too top-down, then they risked not being embedded in practice. Because the tools had stakes attached to their outcomes, they could have been used, but in superficial compliance-based ways and treated as extraordinary events, or worse, those in schools could reject the tools' principles and buffer or resist change completely.

The risk that principals could buffer against the tools was high given that they had been granted new freedoms to be autonomous leaders. The DOE attempted to avoid buffering by balancing the implementation of the accountability tools with structures of support to foster productive use of the tools. For example, a chief architect explained, "I think you've got to empower, but you also have to create structures and you're constantly trying to balance those two things." Yet, this executive commented that, "that's the hardest part of it: how do you continue to require people to have structures through which they can learn and structures through which they're reporting their results?" Autonomy was seen as necessary for allowing the schools to become "autonomous problem-solving units," as described by one former architect, but it also posed complications for the DOE who were faced with the task introducing a new kind of relationship with accountability –

that is to see it as a structure for learning.

One of the ways that the leaders in the DOE attempted to manage this tension was to view their charge of supporting empowered principals as an exercise in instruction. Instead of taking a new management stance where the DOE could simply set targets and punish those who did not meet them, there was an effort to create structures to support principals with developing capabilities to be "autonomous problem-solving" leaders. This point was described by one of the former chief executives,

> I mean Shael (Deputy Chancellor) is not saying 'here's what you should teach and here's how you should teach it.' He's saying, 'here are the standards and you need to figure out how to teach and succeed with these standards and here's a structure within which that can happen and here's a lot of exercises we want everybody to go through to learn about that and that can feel like it's top-down, but he sees it as capacity building for exercising autonomy.

The line that the DOE was trying to walk was balancing requirements with opportunities – similar to what a teacher might do with her students.

A teacher needs her students to learn information – ideally teaching students to think for themselves. The teacher must create scaffolds for students to acquire knowledge and practice independent thinking. The skillful teacher designs learning opportunities for students to do this in authentic ways such that students are able to transfer and apply their newly acquired knowledge and skills in other arenas – a more long term motivation than simply completing an exercise to earn a good grade. This kind of teaching is much more ambitious and challenging than simply lecturing to students, or grading them on what they remembered from reading a text. The DOE was trying to be a skillful teacher; they were trying to figure out how to create useful "exercises" and supports that would enable principals to use the accountability tools with their teachers and staff in ways that would mimic continuous improvement practices and organizational learning. To create useful, instructional structures, the DOE had to undergo a complete transformation of its own.

Similarly important was the need to require that principals attempt to learn the lessons. The DOE was aware that empowered principals would be less inclined to use a recommended tool; it needed to be required. If principals were given a choice as to

whether they should focus on improving the performance of the lowest achieving students, or improving the instructional core and conditions in their schools related to the improvement of teaching and learning, only those principals who possessed the resources (i.e., knowledge, dispositions and talent) would have done so. For example, if "change is scattershot, and primarily bottom-up," according to Tony Alverado, "it will involve only those who perceive a need, and it will yield only small, separate successes" (Liebman & Sabel, 2003, pg. 217).

The architects in NYC sought system level changes – across all schools. They saw the accountability tools to be solutions to what they thought were system level problems (inequality; stagnant, low-quality schools; and wasteful bureaucracy), and used the accountability tools as part of their "systems of solutions to the systems of problems that stood between the schools that existed and the schools they imagined" (Cohen et al, 2014, pg. 28). The combination of the Quality Review, which was intended to provide on-the-ground, school level framework for bottom-up change and the Progress Reports, which were top-down expectations were supposed to work towards balancing top-down requirements with bottom-up learning.

Second, this analysis suggests that the architects attempted to deal with a tension between business tactics and educational professionalism. The form of accountability that they tried to create relying on the Quality Reviews and Progress Reports was meant to blend two opposing forms of accountability that were tied to the standards-movement in the 1990s: lay accountability and professional accountability. According to Mehta (2013), these two different forms of accountability operate on different assumptions about the agenda of reform and the role that motivation and professional knowledge should play in realizing changes.

Lay accountability, according to Mehta, focuses on accountability for results and assumes that strict incentives (read sanctions) will motivate educators to change. NCLB was based largely on this view. By contrast, Smith & O'Day presented a version of expert professionalism that suggested top-down guidance in the form of curricula frameworks based on research and evidence about what works would promote productive changes in practice. Expert professionalism assumed that by making excellent teaching transparent and increasing teachers' knowledge, there would be marked improvement in

student achievement. According to Smith and O'Day, "state of the art examinations based on well-designed curricula frameworks could help encourage instruction toward higher-ordered learning goals" (in Mehta, 2013, pg. 216). Both lay and expert professionalism saw a need for assessments and accountability, but they had different rationales for their use.

In NYC, the architects tried to walk a line between these two opposing views on accountability. However, this was not always perceived in this way – in part because many of the architects had come from different sectors – mainly law and business and were seen as outsiders to many educators. One architect explained how their intentions were misunderstood. As a result, they faced resistance from those who held open animosity toward corporate influence on education. He explained,

> What's interesting about this is that, if you talk to charter people…the language you hear them using and even the criticisms that they had of some of the things New York was doing sounded very, very much like the sort of more thoughtful, professionally oriented people like Linda Darling-Hammond or others so charter schools, have this deep sense of commitment to instruction and things like that. And when we talked about trying to utilize a charter school model, we meant stuff like that – the way that they created these structures in a lot of transparency and a lot of focus, no so different from the effective schools, but what people heard was the corporate stuff and the privatization stuff.

The architects wanted to provide ways for schools to become more like learning organizations, but they also wanted them to be educational learning organizations focused on improving instruction and making those efforts public and the knowledge for teaching transparent within a school. When the architects refer to charter schools, it's not *all* charter schools, but basing on the instructional systems that networks like Achievement First was able to construct within their schools (Rosenberg, 2012). The architects had to contend with people's reactions and misunderstandings. At times, this was made more difficult by the national reform landscape that positioned business and education as opposing forces in reforms. The fact that many of the people leading the reforms in NYC, including the Mayor, were closely associated with business and groups external to education was something that influenced the public's and educator's reactions to the

reforms.

Lastly this analysis points to the challenges of balancing incremental system reform with bold and rapid changes. The architects, in creating, their blueprint for the tools understood that the tools were not perfect when they started implementing them, but went ahead – in part because they were a mission-driven group seeking to radically reinvent the system with a deadline. When Mayor Bloomberg was initially elected, no one could have anticipated that they would have eleven years (especially since there were term limits; Mayor Bloomberg fought to have the law changed so he could have a third term). No one knew in 2007 that they would have seven more years to work on implementing the reforms. As such, they designed the accountability tools specifically to evolve.

The evolutionary process influenced the tools' designs, but the changes that were made were only to "structure conditions among schools, designs, organizations, and environments in order to increase the potential for effectiveness" (Cohen et al, 2014, pg. 61). The feedback and listening tours, the research the DOE conducted on tools in order to make changes was to better align the tools with the core strategy. The outcomes they wanted were to change student achievement and system's culture. Both of these things usually happen in shifts – over time. In NYC, they attempted to change culture – quickly with bold and dynamic changes.

I would have suggested that this was perhaps something only NYC could attempt, but one of the deputy chancellors suggested otherwise. He explained,

> What I think what made it credible in that we are the largest system in the nation and every system should have, in my opinion, a way to look at it schools in a comprehensive way and in a way that is consistent across schools – that standards are about and in a way that I think creates an engine for movement or in a way that helps to helps to capture where your talent is. So it was nothing special about it happening in New York. You know what I'm saying? Because we are able to do it in such a large system, it only, in my opinion, says to smaller systems – smaller districts – you can do this to and you should be doing this because what it does is it sparks conversations and it helps people to look at themselves and to look at your school respectively and your system overall and begin to plan out

from what you see and understand and I think that's a wonderful thing. It's nothing special about New York City other than that we had the courage to do it. We had the wherewithal to do it. We understood and still continue to understand improvement is a process, that it is not going to happen unless you engineer some way of moving things. We've use the progress reports; we've used the quality review; we've used professional development; we've used the teacher evaluations system which is now creating the need for all kinds of support for teachers and principals, but all of that if done well and done right will benefit students. It's nothing special. It's nothing special. We did it that's all. We had the nerve to do it.

The architects "had the nerve to do it," but they also had the resources and the support from a Mayor, who had authority and a desire to completely disrupt and reconstruct the system. The importance of this combination of factors – power, financial resources, human capital, access to research and best practices – shielded from political influence – cannot be understated. This is not to say that NYC had it easy; by no stretch of imagine is that the point. What they attempted to do was tremendously challenging, but this combination of factors definitely made it a little easier.

For researchers, this analysis suggests the importance of extending the descriptions of the Quality Reviews and whether this novel accountability tool helps those schools that do not have access to a networked community, or instructional system develop or expand capabilities to align internal organizational conditions and instructional improvement with external demands for improvement. Future studies could investigate at the school level whether and how schools develop organizational learning and continuous improvement practices in response to accountability pressures in NYC and beyond. Several districts in the U.S. have adopted a performance management approach to district level support. There is much to be learned from comparisons to NYC's approach and the ways in which other districts have approached the design puzzle. As more districts attempt to create useful score cards and accountability metrics, what can we learn about balancing accountability and organizational learning at the school and system level? Finally, this paper foregrounded accountability reforms as a window into the system level reforms that occurred under Mayor Bloomberg. Subsequent analyses and reports could aim to capture better the rich story of system level

overhaul in New York City under Mayor Bloomberg.

When Mayor Bloomberg left office in 2013, the new administration changed the consequences and reduced pressure on schools for results. Did this change how the tools were used? To what extent do people focus use accountability tools as improvement guides, if there are no stakes attached? Knowledge from a study that considered could provide very useful information for districts and states seeking to balance support and pressure for improved school performance and better educational experiences for students.

References

Bisei, M. (1981). Review of *Theory Z: How American Business Can Meet the Japanese Challenge* in *Business Horizons*. Vol. 24, issue 6. Pg. 82-83.

Bloom, H.S., Unterman, R. (2012). Sustained Positive Effects on Graduation Rates Produced by New York City's Small Public High Schools of Choice. MDRC Policy Brief, January 2012. Retrieved: http://www.mdrc.org/sites/default/files/policybrief_34.pdf

Childress, S. Higgins, M., Isimaru, A., Takahaski, S. (2010) Managing for results at the New York City Department of Education. In Education Reform in New York City Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Harvard Education Press. Cambridge, MA.

Cohen, D.K., Peurach, D., Glazer, J., Gates, K., Goldin, S. (2014). Improvement by design: The promise of better schools. The University of Chicago Press: Chicago, IL.

Cohen, D.K., Moffit, S.L. (2009). The ordeal of equality: Did federal regulation fix the schools? Harvard Education Press. Cambridge, MA.

Cohen, D.K., Peurach, D., Glazer, J., Gates, K., Goldin, S. (2014). Improvement by design: The promise of better schools. The University of Chicago Press: Chicago, IL.

Cohen, D.K, Spillane, J.P. (1992). Policy and Practice: The relations between governance and instruction. In G. Grant (ed). Review of Research in Education. 18: 3-49. Washington D.C.: American Educational Research Association.

Daft, R.L., Weick, K.E. (1984). Toward a model of organizations as interpretation systems. *Academy of Management Review*. 9 (2), pg. 284-295.

Edmonson, A., Moingeon, B. (1998). From organizational learning to learning organization. *Management Learning*. Vol., 29, no. 1, pg. 5-20.

Elmore, R.F. (2003). The New Accountability: High schools and high-stakes testing. RoutledgeFalmer: New York, NY.

Elmore, R. F. (2004a). School Reform from the Inside Out: Policy, practice, and performance. Harvard Education Press. Cambridge, MA.

Elmore, R.F. (2004b). Conclusion: The problem of stakes performance-based accountability systems. In Redesigning Accountability Systems for Education. Susan H. Fuhrman and Richard F. Elmore (eds). Teachers College Press. New York NY.

Foley, E.,Mishook, J., Thompson, J., Kubiak, M., Supovitz, J., & Rhude-Faust, M. (2008). Beyond test scores: Leading indicators for education. Providence, RI: Brown University, Annenberg Institute for School Reform.

Fullan, M., Boyle, A. (2014). Big-city school reform: Lessons from New York, Toronto, and London. Teachers College Press: New York, NY.

Gyurko, J., Henig, J.R. (2010) Strong vision, learning by doing, or the politics of muddling through? Katrina E. Bulkley, Jeffrey Henig and Henry Levin (eds). *Between Public and Private: Politics, governance, and the new portfolio models for urban school reform*. Harvard Education Press: Cambridge, MA.

Hemphill, C. (2010). Managing by the Numbers: Empowerment and accountability in New York City Schools. Center for the New York City Affairs at the New School. June 2012.

Hemphill, C., Nauer, K., White, A., Jacobs, T. (2013). Building better schools: How the next mayor can prepare New York's students for college and careers. Education Funders Research Initiative. November 2013.

Honig, M.I. (2006). Building Policy from Practice: Implementation as Organizational Learning. In Meredith I. Honig (ed). New Directions in Education Policy Implementation: Confronting Complexity. State University of New York Press: Albany, NY. Pg. 125-148.

Honig, M. I., DeArmond, M. (2010). Where's the Management in Portfolio Management? In Katrina Buckley, Jeffrey Henig, Henry Levin (eds). *Between Public and Private: Politics, Governance, and the New Portfolio Models for Urban School Reform*. Harvard Education Press: Cambridge, MA. Pg. 195-216.

Klein, J., Rhee, M., Gorman, P., Huberman, R. et al. (2010). How to fix our schools: A manifesto by Joel Klein, Michelle Rhee and other education leaders. *Washington Post.* Sunday, October 10, 2010.

Klein, J. (2014). Lessons of Hope: How to fix our schools. Harper Collins: New York, NY.

Lehigh, S. Joel Klein's advice for Mass. Education Reform. *The Boston Globe.* December 17, 2014.

Liebman, J.S., Sabel, C.F. (2003). A public laboratory Dewey barely imagined: The emerging model of school governance and legal reform. *Review of Law and Social Change.* Vol. 28, pg. 184-303.

Maxwell, J.A. (2005). *Qualitative Research Design: An Interactive Approach*. Sage Publications (2nd ed.). Thousand Oaks, CA.

Mehta, J. (2013). The Allure of Order: High hopes, dashed expectations and the troubled quest to remake American schooling. Oxford University Press: London.

McDonald, J.P. (2014) American school reform: What works, what fails, and why. University of Chicago Press: Chicago, IL.

McDonald, J.P., Klein, E., Riordan, M. (2009). Going to scale with New School Designs: Reinventing high school. Teachers College Press: New York, NY.

McLaughlin, M.W. (2006). Implementation Research in Education: Lessons Learned, Lingering Questions and New Opportunities. In Meredith I. Honig (ed). New Directions in Education Policy Implementation: Confronting Complexity. State University of New York Press: Albany, NY. Pg. 209-228.

Miles, M., & Huberman, M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher, 13*, 20–30.

Nadelstern, E. (2013). 10 Lessons from New York City Schools: What really works to improve education. Teachers College Press: New York, NY.

Newmann, F.M., King, M.B., Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*. 67, 1; pg. 41-65.

New York City Department of Education. (2008). Children First: A Bold, Common Sense Plan to Create Great Schools for All New York City Children. Updated: 2008-2009 School Year.

New York City Department of Education. (2012). Progress Reports for New York City
    Public Schools Powerpoint prepared by the Office of Performance and
    Accountability in Fall 2012.

New York City Department of Education. (2013). What's Next for School Accountability
    in New York City? Released November 2013.

New York City Department of Education Quality Review Rubric (2015).   Retrieved:
    http://schools.nyc.gov/NR/rdonlyres/8C11A001-7E78-469D-996F-
    B0C3703CEA81/0/201314QualityReviewRubric.pdf May 21, 2015

NYC DOE Quality Review 2010-11 Academic Quality Division of Performance and
    Accountability, September/October 2010

O'Day, J. A. (2002). Complexity, accountability, and school improvement. In
    Redesigning Accountability Systems for Education. Susan H. Fuhrman and
    Richard F. Elmore (eds). Teachers College Press. New York, NY.

O'Day, J., Bitter, C. (2010). Improving Instruction in New York City Schools: An
    Evolving Strategy. In Jennifer O'Day, Catherine Bitter, Louis Gomez (eds).
    Education Reform in New York City. Harvard Education Press. Cambridge, MA.

Ouchi, W.G. (2008). Making schools work: A revolutionary plan to get your children the
    education they need. Simon & Schuster: New York, NY.

Ouchi, W.G. (2013). The Secret of TSL (Total Student Load): The Revolutionary
    Discovery that Raises School Performance. Simon & Schuster: New York, NY.

Peurach, D. (2011). Seeing complexity in public education: Problems, possibilities and
    success for all. Oxford University Press.

Phillips, A. (2012). City gives $5.7 million to principals in bonuses. WNYC February
    15, 2012. Retrieved:
    http://www.wnyc.org/story/301638-city-gives-nearly-6   million-to-principals-
    in-bonuses/ May 21, 2015.

Poole, M. S. (2004). Central issues in the study of change and innovation. In M. S. Poole
    & A. H. Van de Ven, (Eds.), Handbook of organizational change and innovation
    (pp. 3-31). New York: Oxford University Press.

Rosenberg, S. (2012). Organizing schools for quality in education: Individualistic and
    systemic approaches to teacher quality. Dissertation: University of Michigan.

Senge, P.M. (1990). *The fifth discipline: The Art and Practice of the Learning
    Organization*. New York: Currency Doubleday.

Shanker, A. (1993). A tribute to Albert Shanker. The Pew Forum on Education Reform.
    Retrieved: http://nevadajournal.com/assets/uploads/2014/04/shankerpew-0.pdf

Weiss, R.S. (1994). *Learning from Strangers: The Art and Method of Qualitative
    Interview Studies*. The Free Press: New York, NY.

Yin, R.K. (1984). Case study research: Design and methods. Sage Publications: Beverly
    Hills, CA.

APPENDIX A: NYC Progress Report

NYC Department of Education
Dennis M. Walcott, Chancellor

Progress Report 2011-12

Aviation Career & Technical Education High School

PRINCIPAL: Deno Charalambous

DBN: 24Q610
ENROLLMENT: 2201
SCHOOL TYPE: High School
PEER INDEX*: 2.97

*See p. 7 for more details on Peer Index.

PROGRESS REPORT

A

OVERALL SCORE **76.4** out of 100

OVERALL PERCENTILE **82**

This school's overall score is greater than or equal to that of 82 percent of high schools.

For high schools, grades are based on cut scores determined prior to the release of the Progress Report. Further, schools with a four year graduation rate in the top third citywide cannot receive a grade lower than a C. Schools in their first year, without a graduating class or in phase out receive a report with no grade or score.

Progress Report Grades - High School

GRADE	SCORE RANGE	% OF SCHOOLS
A	70.0 or higher	35% of schools
B	58.0 - 69.9	37% of schools
C	47.0 - 57.9	20% of schools
D	40.0 - 46.9	5% of schools
F	39.9 or lower	3% of schools

QUALITY REVIEW

WD

Well Developed (2009-10)

The rating is based on three major categories of school performance: instruction that prepares students for college and careers, school organization and management, and quality of the learning environment.

A school that receives a Well Developed rating earned the highest grade for highly effective teaching and learning practice, strategic school management, and an excellent quality learning environment. For more information, see:

http://schools.nyc.gov/Accountability/tools/review

Overview Each school's Progress Report (1) measures student year-to-year progress, (2) compares the school to peer schools, and (3) rewards success in moving all children forward, especially children with the greatest needs.

CATEGORY	SCORE		GRADE	DESCRIPTION
Student Progress	**39.5** out of 55		A	Student Progress measures the annual progress students make toward meeting the state's graduation requirements by earning course credits and passing state Regents exams.
Student Performance	**14.9** out of 20		A	Student Performance measures how many students graduated within 4 and 6 years of starting high school, and the types of diplomas they earned.
School Environment	**12.1** out of 15		A	School Environment measures student attendance and a survey of the school community rating academic expectations, safety and respect, communication, and engagement.
College and Career Readiness	**7.0** out of 10		A	College and Career Readiness measures how well students are prepared for life after high school on the basis of passing advanced courses, meeting English and math standards, and enrolling in a post-secondary institution.
Closing the Achievement Gap	**2.9** (16 max)			Schools receive additional credit for exceptional graduation and college/career readiness outcomes of students with disabilities, English Language Learners, and students who enter high school at a low performance level.
Overall Score	**76.4** out of 100		A	The overall grade is based on the total of all scores above. Category scores may not add up to total score because of rounding.

Source: NYC Department of Education, 2012. Retrieved:
http://schools.nyc.gov/OA/SchoolReports/2011-12/Progress_Report_2012_HS_Q610.pdf

APPENDIX B: Quality Review Rubric Big Ideas

**Quality Review 2013-14: Big Ideas by Indicator and Sub-indicator**

Instructional Core			School Culture	
Indicator 1.1* **Rigorous, engaging and coherent curricula aligned to CCLS** a) Curricula align to CCLS and/or content standards and instructional shifts b) Rigorous habits and higher order skills for all c) Planning and revising to ensure access to curricula and cognitive engagement for all	**Indicator 1.2*** **Research-based, effective instruction that yields high quality student work** a) Shared beliefs informed by the Danielson framework are aligned to pedagogy and curricula b) Teaching strategies provide multiple entry points that engage all learners c) High levels of student thinking and participation culminate in meaningful work products	**Indicator 2.2*** **Curricula-aligned assessment practices that inform instruction** a) Curricula-aligned assessment practices and grading policies provide actionable feedback b) Common assessment analysis drives curricular and instructional adjustments c) Checks for understanding and student self-assessment lead to effective lesson adjustments	**Indicator 1.4** **Structures for positive learning environment, inclusive culture, and student success** a) Safe and inclusive school culture b) School coordinates social-emotional learning, attendance supports, and youth development for academic success c) Aligned professional development and supports for adoption of effective academic and personal behaviors	**Indicator 3.4** **Establishing a culture of learning that communicates high expectations with supports** a) Communication of high expectations to staff, including training and a system of accountability b) School leaders and staff communicate and support high expectations for college and career readiness to families c) Staff communicate and support high expectations to students
Systems for Improvement				
Indicator 1.3* **Aligned use of resources to support instructional goals that meet students' needs** a) Resource use aligns to instructional goals b) Use of time improves instruction and challenges students c) Student programs align to teacher talent and support access to learning that leads to college and career readiness	**Indicator 3.1** **School-level theory of action and goals shared by the school community** a) School-level goals and theory of action are tracked for progress and accelerate student learning b) Data driven needs assessments inform school level goals, action plans and professional development c) Community involvement in setting school level goals and action plans	**Indicator 4.1*** **Support and evaluation of teachers through feedback using the Danielson framework and analysis of learning outcomes** a) Teacher growth supported by effective feedback and next steps from observations and data b) Feedback to teachers supports development, offers trends and next steps using the Danielson framework c) Data informed professional development and teacher/administrator performance based decisions and succession planning	**Indicator 4.2** **Teacher teams engaged in collaborative practice using the inquiry approach to improve classroom practice** a) Teacher teams engage in collaborative inquiry that supports goals and strengthens teacher capacity b) Student work/data analysis within teams improves curricula, teaching and learning c) Embedded distributed leadership structures influence key decisions	**Indicator 5.1** **Regularly evaluate school level decisions with a focus on the CCLS** a) Evaluating and adjusting curricular and instructional practices based on student needs b) Evaluating school culture and expectations to make adjustments c) Evaluating and adjusting use of resources, teacher team effectiveness and professional development

Source: NYC DOE, 2014 Office of Performance and Assessment, Quality Review

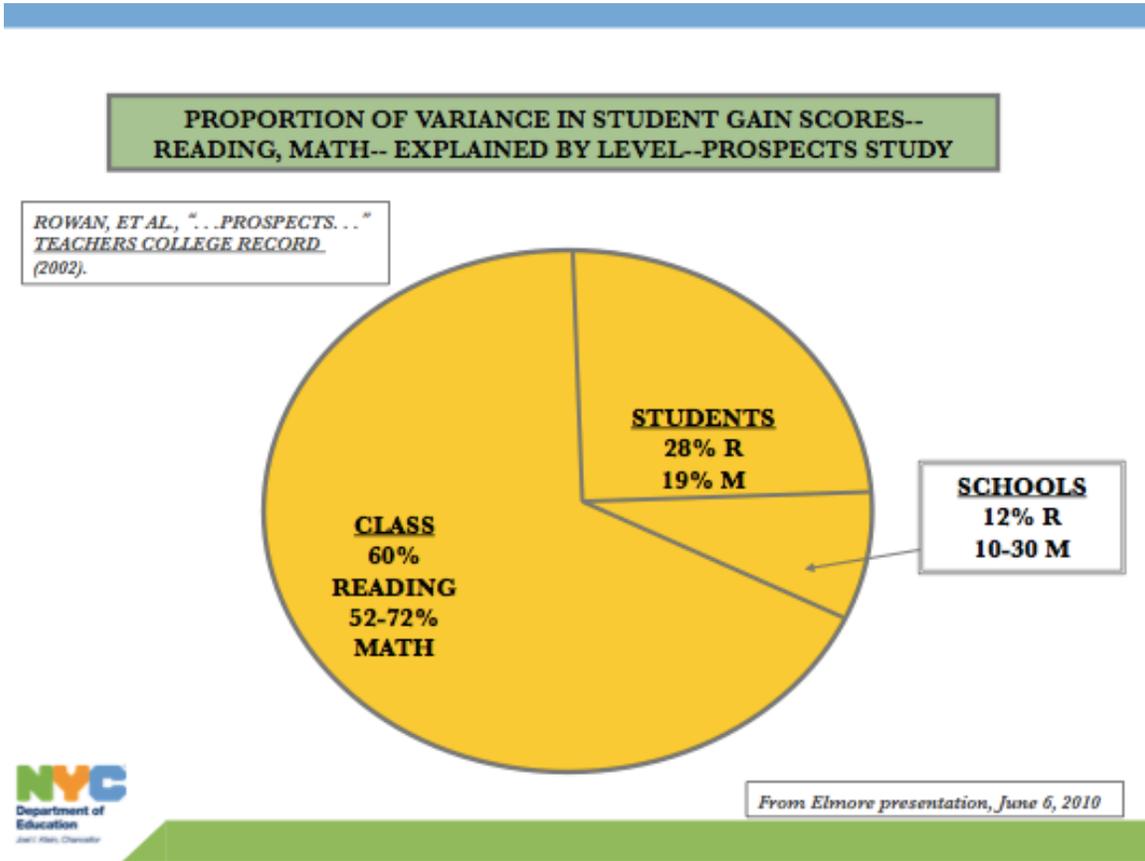APPENDIX C: NYC DOE Progress Report Overview

PROGRESS REPORT OVERVIEW

Progress, Performance, School Environment, and High School/College & Career Readiness scores based on comparison to peer schools (75%) and City (25%).
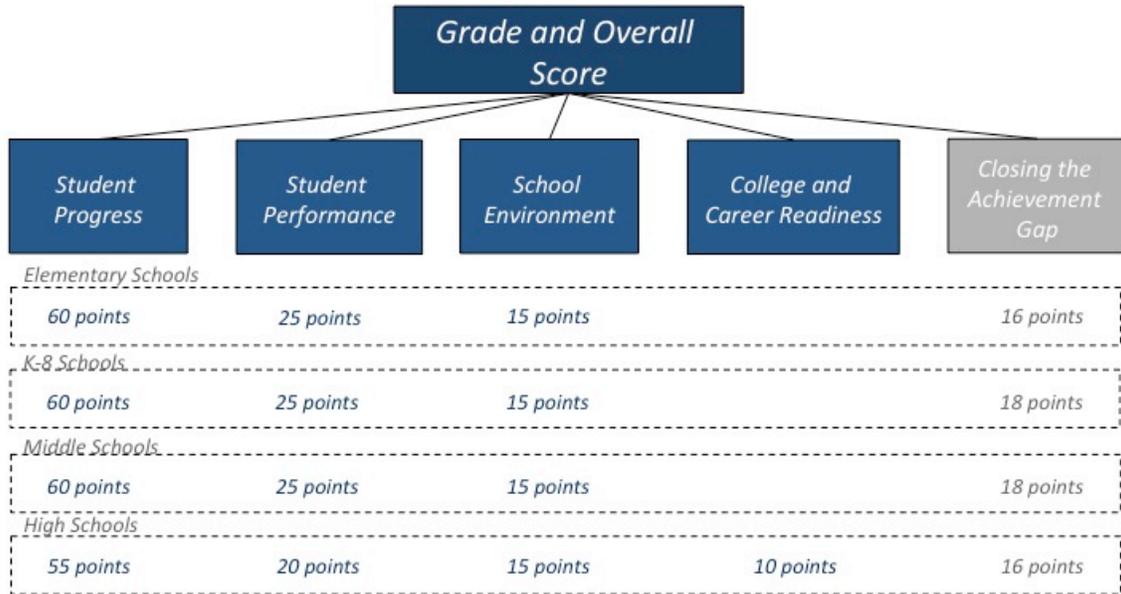
Source: NYC DOE Presentation: Progress Reports for New York City Schools, March 2014, slide 9.

APPENDIX D: NYC DOE Motivating shift to instructional coherence
in QR rubric (Fall 2010)

PROPORTION OF VARIANCE IN STUDENT GAIN SCORES--
READING, MATH-- EXPLAINED BY LEVEL--PROSPECTS STUDY

ROWAN, ET AL., "...PROSPECTS..."
TEACHERS COLLEGE RECORD
(2002).

STUDENTS
28% R
19% M

SCHOOLS
12% R
10-30 M

CLASS
60%
READING
52-72%
MATH

From Elmore presentation, June 6, 2010

Source: NYC DOE Quality Review 2010-11 Academic Quality Division of Performance and

Accountability, September/October 2010, slide 6.

CHAPTER IV

SCHOOL SUPPORT AND ACCOUNTABILITY: NETWORKS' USE OF THE
PROGRESS REPORTS AND QUALITY REVIEWS IN NEW YORK CITY

"The central dilemma of growth is reconciling the demands of learning with the demands of monitoring. By economic learning I mean acquiring the knowledge to make and do things valued in the markets. This of course supposes unlearning knowledge that is not so valued."

- Charles Sabel, 1993

Over the last decade, in several cities across the country, including Los Angeles, New York and places in between, mayors have taken over control of the public schools. Seen as a response to mounting discontent with large bureaucracy that was believed to stymie innovation, these strategies were propelled by forces seeking to put the central office in the "distinctly new role as strategic manager of change" (Bulkley, 2010, pg. 5).

When Mayor Bloomberg took over of the public schools in New York City in 2002, he made two strategic moves to change 'business as usual.' The first changed who held power in the district: Replacing the Board of Education with a newly designed Department of Education. He eliminated thirty-two community school boards that had ruled local politics since 1969 and replaced them with ten administrative regions. In addition, he dismantled the elected school board in favor of an advisory group called the Panel on Education Policy of which he appointed eight of the thirteen seats.

The second was a symbolic break from the past: He ordered that the Department of Education offices be moved from Borough Hall Brooklyn to the historic Tweed Courthouse, located near Manhattan's City Hall grounds - a move that brought central office closer to the mayor who had won his first political post by committing to "fix public education." The move was symbolic and strategic. Tweed Courthouse's maximum capacity was six hundred people. To make the move to Manhattan, central office had to become a leaner organization, which fit better with Mayor Bloomberg's

vision for the Department of Education (Nadelstern, 2012). These were unmistakable departures from NYC's past.

Paul Hill (2010) described Mayor Bloomberg's strategy as decentralization - a seemingly contradictory description given that the mayor seized control from 32 local boards. What Hill (2010) argued was that Bloomberg's brand of decentralization was as it is practiced in business.

Hill wrote, "Though in education, decentralization has often been an effort to encourage initiative at the school level without changing the constraints imposed by central offices and unions, in business, decentralization is something else" (pg. 4). In business, decentralization strengthens the top and bottom of an organization, but weakens the middle.

According to Hill, the middle is mainly comprised of "middle managers" – people responsible for enforcing rules, controlling hiring, or routinizing operations; middle management stands between the CEO and the people doing the day-to-day work. The value of the middle manager is typically bureaucratic - not technical (Hill, 2010). In education, the analog of the middle manager is the superintendent who uses command and control tactics to push schools toward standard, routinized procedures focusing more on principals' compliance with rules than the development of professional expertise.

Mayor Bloomberg's early moves definitely strengthened the top (central office), but he also made several moves to strengthen the bottom - namely, principals. In what was referred to as an authority-accountability exchange, Mayor Bloomberg devolved power to principals to act as CEOs with authority to make decisions regarding budgets, curriculum, instruction, and staff (Childress et al, 2010; Wohlstetter et al., 2012). In exchange, principals were held accountable for hitting specific targets and goals as outlined by the Progress Reports and Quality Reviews (O'Day et al, 2010; Childress et al, 2010; Wohlstetter et al, 2012). According to Klein, "Without a great leader, a school - like any other organization - doesn't work well. To make schools efficient and effective, principals needed to be empowered and supported from above to get classroom teachers up to snuff and get rid of those who couldn't hack it" (Klein, 2014, pg. 23).

Klein's sentiment reflects more than brash disdain for incompetent teachers. Rather, it signals a set of ideas that were central to Chancellor Klein's and Mayor

Bloomberg's approach to "fix" the system. More than a decade later, the ideas driving their reforms are now recognized and referred to as a Portfolio Management Model (PMM). PMM is a fairly broad based term that refers to a shift from a top-down, standard delivery model towards a district that provides various school options, including charters, semi-autonomous, conventional configurations. Klein explained that the aim of the strategy was "…not a great school system, but a system comprised of great schools, and we made it clear that the individual school would be the key link in our organizational structure" (Klein, 2014, pg. 48).

A key feature of PMM is accountability for performance – no matter how they are governed – under PMM, schools are held accountable for performance. Accountability involves closing schools that cannot hack it and opening new ones under the distinct expectations for performance. During a speech at an education conference in Washington D.C. in 2009, Mayor Bloomberg stated, "Our goal is to turn around the lowest-performing 10 percent of city schools over the next four years by closing them down and bringing in new leadership and holding everyone accountable for success" (Cited in Hemphill, 2010, pg. 46). According to Hemphill, Mayor Bloomberg said, "Secretary Duncan has challenged states to turn around their lowest-performing 5 percent of schools. Arne, we'll see your 5 percent and we're going to double it." This public promise was a commitment to close nearly forty schools per year by the end of the Mayor's term in 2013.

According to Paul Hill and his colleagues (2009), "A district fully committed to portfolio management would hold all schools, educators, and providers, no matter whether they are district employees or outsiders, equally accountable for performance defined by student achievement and attainment, abandoning less productive schools and arrangements, and sustaining or expanding more productive ones" (Hill, et al, 2009, pg. 7). Contrasting traditional, bureaucratic public-school management, portfolio management eliminates protections for schools or employees; staying in business is contingent upon performance (Hill et al, 2009).

Given this strategy, Mayor Bloomberg sought to weaken the middle. He eliminated the balkanized superintendent structure (that employed nearly six thousand bureaucrats) and stripped superintendents of their traditional responsibilities. Networks

were created to provide non-supervisory support to schools. By 2010, all principals in the city choose one of 60 Children First Networks as a partner for instructional and operational support. Networks, too, were subjected to strict accountability for performance and operated in what can be described as a public-sector market.

Because principals could select their network of choice, these networks functioned in a market for their services. If a principal was dissatisfied with her network, she could affiliate with a different network. This created a market-place in which networks competed to serve schools and satisfy principals. Yet, this was not simply market-based rules; as is the case for schools under PMM models, networks, too, were regulated by the DOE's central office and subjected to similar performance standards that schools, principals, and teachers had to meet. Networks were evaluated annually using a Network Performance Management (NPM) rating system, which held networks accountable for their schools' outcomes on the Progress Reports and Quality Reviews, and results from a principal satisfaction survey. If too many schools left a network and the number of schools in a network dropped below eighteen or fewer, or if a network persistently underperformed on NPM, then the network was disbanded.

When Mayor Bloomberg and Chancellor Klein adopted a portfolio model approach and restructured NYC's school system, they did so for all schools regardless of prior performance. This is distinct from other districts where autonomy was only awarded to higher performing schools that had proven successes with student achievement. The logic in NYC was that autonomy was a necessary condition for improving student outcomes. Autonomy was off set by a new performance-based accountability system that relied on Progress Reports and Quality Reviews as the primary metrics attached to consequences. Bloomberg and Klein shifted power and expectations for performance, which created novel demands on support relationships to achieve unprecedented results. There is plenty to explore in such unconventional circumstances. Yet, what I set out to understand was how networks, working under unfamiliar pressures for performance, used the Progress Reports and Quality Reviews in their efforts to support schools.

The networks were in the unique position of having to navigate the space in between schools and central office; they needed to be able to help schools with their

challenges and find ways to serve principals who needed to develop capabilities as lead problem-solvers. Simultaneously, networks had to find ways to meet central office's expectations that everyone in the system "put their best practice forward." This meant that networks needed to be experts in the accountability tools and be able to guide others in using them. They had to be responsive both to central office (their Tweed masters) and their school clients.

And it was not easy work because under Mayor Bloomberg and Chancellor Klein, no one, especially principals, was going to be let off the hook for poor performance. The networks were created to provide supports to principals who needed to move student learning – no matter what. By design, networks were external support providers charged with helping principals develop capabilities for continuous improvement practices – regardless of whether the principals were leading schools that were well positioned to take on the tasks of continuously improving teaching and learning. How did the networks do this unusual work?

In the sections that follow, I answer that question in stages. First, I provide background on NYC's Children First Networks and present an overview of the framework that I used to explore these uncommon support teams' use of the Progress Reports and Quality Reviews. Then, I draw on an image of the networks as brokers working under unconventional circumstances to target support and adapt to meet schools' diverse needs. This research extends and contradicts past research that suggested external school support often did little to improve educator or organizational performance as it suffered from a lack of focus, weak engagement or vague goals (i.e., Finnigan, Bitter, and O'Day, 2009; Ikemoto & Honig, 2008).

General background on Networks in New York City

From 2010-2014, all schools in New York City, including public and charter schools, received instructional and operational support from a Network team. Across the city, there were approximately 60 Networks referred to as Children First Networks. Each Network served 25-30 schools and were grouped into a Cluster. In 2013, there were five Clusters supporting approximately 11 networks each. Cluster teams worked closely with

the Department's central leadership to oversee and support Network teams (See Appendix E).

Network teams were comprised of approximately fifteen people who were experienced educators and school leaders. Most team members possessed expertise in particular areas such as instruction, pedagogy, special education, budgets, technology, or student safety. This kind of cross-functional team model was distinctly different from the silos of support that existed under the prior regional structure.

Networks were not geographically based; it was not uncommon for Network teams to travel across all five boroughs to provide onsite support. In 2013, only four of the more than fifty Networks served schools from one borough; not a single network was comprised of schools from only one district (Parthenon, 2013). The self-organized Networks were allowed to brand themselves around a particular instructional model or philosophy; some were geared to serve a particular type of school (I.e., specialized high schools or transfer schools serving more overaged students). Principals selected to partner with a network based on preference, targeted needs of students and communities, or personal relationships.

If principals were dissatisfied with their network, there was an open enrollment period during the spring when principals could affiliate with a different network. Ninety-percent of schools remained with their networks suggesting that for the most part, principals were satisfied with their networks (Parthenon, 2013). The DOE contracted with KPMG to manage its annual principal satisfaction survey. The survey was intended to help the DOE improve the quality of the services and supports provided to schools based on principals' feedback. In 2013, 84% of the principals who took the survey reported that they were satisfied with the "overall quality of support provided by the network."

Until 2013, networks were evaluated annually under the DOE's Network Performance Management (NPM) system. NPM was a ranking system that evaluated networks using a set of weighted factors that included schools' Progress Reports, Quality Reviews, results from the principal satisfaction survey and a qualitative assessment of networks' functional areas. The qualitative evaluation tools were created by the DOE's Office of School Support to assess how well a network team "holistically and

collaboratively supports all of their schools and continuously develops their competencies to implement the Citywide Instructional Expectations and other expectations and work flows outlined in the Functional Framework" (DOE, 2013). Persistently ineffective Networks were disbanded. Between 2011-2013, the DOE closed eight networks due to poor performance and launched six new ones in their place (Parthenon, 2013).

Before rolling out the network structure and Portfolio Management Model of governance for all schools in the city, the DOE spent two years piloting the model in an experiment called the Autonomy Zone. The Autonomy Zone was established in 2004 with a cohort of twenty-nine schools. Schools in the Autonomy Zone were allowed to organize themselves to solve issues related to the coordination of instruction and organizational conditions; these schools were granted as much freedom as law would permit to decide curriculum, professional development, staffing and budgets – a release from bureaucratic red tape that bound schools to central office in compliance based relationships. Instead, principals in the Autonomy Zone were excused from attending mandatory DOE meetings and writing reports for supervisors.

Support for schools in the Autonomy Zone looked different, too. Rather than get support from the regional structure like the other schools in the city, schools in the Autonomy Zone received support from cross-functional teams that helped with a broad range of issues including instructional coaching, budgets, data management, and school attendance. As the pilot scaled-up, more schools entered the Zone and more network teams were established. In 2008, the Autonomy Zone got a new name: the School Support Organization (SSO). The SSO served more than six hundred schools.

Eric Nadelstern, the former Deputy Chancellor for the Division of School Support from 2009-2011 and first Chief Academic Officer for the Autonomy Zone was a staunch proponent of principal autonomy. His beliefs were supported by experience as a pioneer of reform in the city during the 1980s. He explained why Autonomy Zone made sense, "We were experimenting with the idea that if we protected schools from outside interference and did little else but held them accountable for outcomes and allowed them to make important decisions would they get better results?" The Autonomy Zone pilot provided an answer: Yes. Autonomy was good for school improvement. Nadelstern

explained, "the autonomy zone schools not only outperformed the system, but they also outperformed their own previous performance."

Nadelstern championed autonomy as a strategy as long as the DOE could get accountability and support right. Getting the balance right, according to the original architects, resembled Richard Elmore's reciprocal accountability such that principals needed to be provided supports to develop the capabilities needed for more ambitious results; central office needed to supply the supports, if it was going to expect different results.

Reciprocal accountability can be expressed simply as, "For every increment of performance I demand from you, I have an equal responsibility to provide you with the capacity to meet that expectation. Likewise for every investment you make in my skill and knowledge, I have a reciprocal responsibility to demonstrate some new increment in performance" (Elmore, 2000). Such a concept was foreign under NYC's old configuration.

Before becoming Chief Academic of the Autonomy Zone, Eric Nadelstern served forty years as a teacher and principal in the system; during those years he worked for fourteen different chancellors. Of this time, Nadelstern commented, "not one of those individuals was replaced because kids were not learning, even though, during most of that time, only fifty-percent of students were graduating from high school." Another illuminating example comes from his book, *10 Lessons from New York City Schools:*

> When I arrived at the Bronx Superintendent's Office, there was a staff of 120 serving 20 high schools…the office was typically organized into functional units: instruction, guidance, special education, technology, human services, and operations. When asked which schools they were responsible for, every staff member said all of them. But when asked what exactly each was accountable for, the question invariably resulted in silence and quizzical looks. Everyone was responsible for everything, but no one was accountable for anything (pg. 60).

Under the old structure, support for school improvement was not subjected to any form of accountability, as described by another senior official: "…from top down, there was absolutely no accountability for student performance; there was only accountability for getting your paperwork in on time and for making sure you got along with your

immediate supervisor. If you could do those two things, you didn't have to worry about whether or not what you were doing had a positive effect on children."

When the DOE granted autonomy for all schools in 2007, it also changed the support structure – offering schools principals the freedom to choose their support provider. By 2010, all principals had to choose a partner from the nearly sixty Children First Networks. The shift in the governance structure and the new model of school support created unique conditions for schools, support, and central office – the interdependencies at work were significant and created new demands for new kinds of capabilities.

According to Meredith Honig and Michael DeArmond (2010) under a portfolio management model, central office must perform three key functions in order for this approach to provide improved results. The first is that central office needs to cultivate relationships with schools and outside groups to develop a supply of high quality school options. Secondly, according to Honig and DeArmond, central office must develop or manage an accountability system to gauge whether students are learning from the differentiated providers across the system and track and provide tools to support system progress. These authors use New York City's Office of Accountability as an example. They described how the DOE launched four tools to "assess school performance and inform support and accountability decisions," including the implementation of the Progress Reports, Quality Reviews, and two other tools, periodic assessments and the Achievement Reporting and Innovation System (ARIS) (Honig & DeArmond, 2010, pg. 199). Lastly, these authors described that central office must be equipped to close schools that do not meet the performance targets set by central.

Under the new form of governance, principals were seen as primary problem-solvers – responsible for solving any and all of the problems that needed to be solved so that all students were making progress. Although moving students, and changing schools to deliver effective education for all students – especially the lowest third of students is rarely a direct, linear process. There are a host of problems that may impede student learning. Certainly there are factors beyond a school's control: poverty, violence, stress in a child's home, access to books and enriching activities like museums or musical instruments. However, problems also arise that are directly within a school's purview

like whether there is a safe and caring environment for students, the strength of professional collaboration among teachers, effective leadership, and quality of instruction (Byrk et al, 2010). For the first time, all schools, regardless of past performance, or circumstance were being told that they had to provide value to all students, which meant that schools had to develop sets of capabilities that they may not have possessed in the past – and needed to learn these things in a timely way as there was pressure to prove that they could hack it in the newly established portfolio management model run by central office.

One of the original architects reflected on how the reforms under Mayor Bloomberg and Chancellor Klein changed roles, responsibilities and relationships for central office and schools:

> If you want to have a more systemic, pervasive impact on student achievement, then you can't do the things we've always done and expect radically different results. You have to do something different. What we did in New York was to say the legitimate role of the central office is to find the best people we can find to be principals and if you can't find them, train them yourself as in the New York City Leadership Academy. We started our own principal preparation program. Support them, protect them from outside interference, if you can provide them with incentives when they do good work, but most importantly, hold them accountable for the outcomes. And that's different than saying, we're going to tell you what you need to do as a principal and a teacher. If you tell people how to do it, then all you succeed in doing is letting them off the hook because then if it doesn't work, it's not their fault, it's your fault, whereas if you give them the opportunity to be professionals, to innovate, to put their best practice forward, then you're in a position to one, remove the obstacles that keep them from success but ultimately hold them accountable for making sure that their students succeed and if not, giving other people a chance to.

This shifting nature of managing schools and school improvement created unique challenges and opportunities for central office and schools. At the individual school level, the challenges may have been experienced in vastly different ways depending on the resources available to a school. By resources, I mean the talent, norms, culture,

collective knowledge (including past knowledge) and systems for learning that could help, or hinder a school as it tries to adapt to the external demands outlined by the Progress Reports and Quality Reviews. The diversity of challenges and degree to which some schools could meet the expectations, while others could not created another level of challenges for central office because it needed to be able to evaluate schools and hold them accountable.

The networks were in the unique position of having to navigate the space in between schools and central office; they needed to be able to help schools with their challenges and find ways to serve principals who needed to develop capabilities as lead problem-solvers. Simultaneously, networks had to find ways to meet central office's expectations that everyone in the system "put their best practice forward." This meant that networks needed to be experts in the accountability tools and be able to guide others in using them.

Under Mayor Bloomberg and Chancellor Klein, no one, especially principals, was going to be let off the hook for poor performance. The networks were created to provide supports to principals who needed to move student learning – no matter what. By design, networks were external support providers charged with helping principals develop capabilities for continuous improvement practices – regardless of whether they were leading schools that were well positioned to take on the tasks of continuously improving teaching and learning. In the section that follows, I provide an overview of the conceptual framework used for considering this novel support arrangement.

Literature

The Children First Network teams in New York City were charged with helping principals manage school-wide responses to a new kind of bureaucratic pressure coming from central office. Principals were under pressure to make sure *all* students were making progress (lagging indicators) and to develop coherent and coordinated conditions to promote stronger teaching within the schools (leading indicators). The evaluations would be made public with the Progress Reports and Quality Reviews.

By design, this arrangement assumed that schools would understand and be able to respond to the new pressures, and where they struggled to do so, that principals would turn to their network-partner for support. A network team's success depended on its

principals being capable and willing to ask for help. The teams also needed to posses

adequate knowledge and skill to provide targeted supports to meet each school's

particular needs. There are three assumptions made by this design for support. The first

is that schools would be able to respond to external pressures, and in doing so, principals

would be able to specify what kind of support they needed to respond successfully, or be

able to negotiate with and learn from the network partnership. Second, there is an

assumption that external support could bolster school improvement. Lastly, the logic

suggests that combining market and bureaucratic accountabilities would increase the

quality of supports provided to schools. In the sections below, I discuss these core

assumptions and what research tells us about each.

External pressure for performance

Education research presents, largely through negative examples, how low performing

schools, in particular, struggle to respond to external demands for changed performance –

especially when under threat from sanctions. The changes that schools make are often

structural like altering students' schedules to require more after-school tutoring (Marsh &

Ikemoto, 2006), or allocating more resources for tested subjects (Diamond & Spillane,

2004). However, several scholars have raised concerns about negative practices that

schools take up in response to external pressure, which include, but are not limited to

narrowing curriculum, lowering standards or gaming that occurs as individuals try to

cheat the system or avoid punishments (Kane & Staiger, 2002; Linn, 2000; Murnane,

1987).   These kinds of responses do very little to promote stronger coordination of

internal conditions like norms and routines that make more coherent and better

instruction possible and suggest that when schools are expected to determine on their

own how to proceed and develop capabilities under pressure, it requires a herculean effort

that is likely out of reach for schools operating as independent organizations.

There is very little evidence that mediocre or worse schools transform into great

ones in response to external pressure. Schools on the sanction list rarely move and stay

off the list, in part, because in those schools there is "limited attention to the technical

aspects of organizational learning, especially at the schools that had been on the sanction

list the longest" (Finnigan and Daly, 2012, pg. 64). The technical aspects, according to

Finnigan and Daly, include how schools detect and correct errors and how a school acquires, develops and refines ideas and routine practices. Low-performing schools often suffer from lower levels of trust, support, and respect among the staff, which tends to obstruct the flow of information that is needed to make continuous improvement possible (Finnigan & Daly, 2012; O'Day, 2002). Rather, places where there are stable, shared commitments to excellent teaching, high expectations for all students and productive professional norms for collective problem-solving around the instructional core are places that are well positioned to respond to pressure from the outside (Elmore, 2004).

Several studies have highlighted that a school's internal conditions including the formal structures that promote collaboration among teachers focused on improving instruction, professional norms for inquiry, formal structures for collective problem-solving, technical expertise related to instruction, and leadership that cultivates an open, trusting culture for learning – these internal conditions predispose a school to be responsive to external pressure (Finnigan & Daly, 2012; Mintrop & Trujillo, 2007; Elmore, 2004; O'Day, 2002; Newmann, King, & Rigdon, 1997).

Scholars have suggested that these internal conditions promote an *internal* accountability, which influences the extent to which teachers and principals feel responsible to one another to improve student learning and the degree to which they turn to the structures, culture and social norms inside their schools to make decisions; internal accountability can motivate changes in behavior based on shared goals, but the strength of internal accountability varies tremendously from school to school (Elmore, 2004).

Early research that investigated teachers' perceptions of internal and external accountability in New York City confirmed that schools where teachers felt safe and perceived their schools as places with strong internal accountability (processes and culture for learning) outperformed those schools where teachers did not perceive a safe, learning culture (Childress et al, 2010). Evidence suggests that social processing of knowledge, relationships and school culture are key elements that determine a school's responsiveness and adaptability, but these are rarely taken up in policy or accountability reforms (Finnigan & Daly, 2012). Due to an absence of accountability measures that focus on the social processes inside schools, there is limited knowledge about whether new measures that require schools to account for their formal structures for

organizational learning will encourage stronger adaptive practices in schools, or whether such measures will equip principals with the necessary knowledge to identify the kinds of supports they may need to attend to school culture or internal structures and processes more broadly.

External support

Schools and districts turn to external support providers – also referred to as 'brokers' or 'boundary spanners' to provide new ideas, resources, models of practice, or alternative forms of support not available within a traditional district-school relationship (Datnow & Honig, 2008). Past research on external support for school improvement has focused on the scaffolds and resources that brokers provide to support learning through social interactions and collective sensemaking (Honig & Ikemoto, 2008). Rarely do brokers serve as supervisors; rather, they typically reside and function outside of the bureaucratic structure. Given their outsider position, they are principally concerned with providing assistance with reform initiatives at the school level that are directed at changing practices in schools (Honig & Hatch, 2004; Honig, 2006; Ikemoto & Honig, 2008; Finnigan, Bitter, and O'Day, 2009; Honig & DeArmond, 2010).

There is value in looking at brokers and what they do in their relationships with particular schools, as well as, their work across a district. Brokers may be responsible for fostering roughly similar forms of learning across diverse schools. This requires them to travel across boundaries and work within different norms and valued ways of knowing and doing that define school cultures. Yet, brokers must also know how to 'hold steady' so that learning in these various contexts is not haphazard or misdirected. Coburn & Stein (2008) suggest that one way brokers do this is to rely on "boundary objects" or "concrete objects that embody a set of ideas or processes" that serve as a "point of focus" for brokers working across different contexts (pg. 589). One example of a boundary object, according to these authors, is pacing guides. Coaches introduce the same pacing guides to teachers in different schools, and in turn, those teachers may meet at district-wide meetings to discover and learn from how the other schools use the guides. Pacing guides, or boundary objects more broadly, "have the potential to coordinate perspectives

and spur similar forms of learning across multiple communities" (Star and Griesmer, 1989 in Coburn & Stein, 2008, pg. 598).

As go-betweens, brokers do more than simply deliver and transmit information; rather, they may participate in developing shared plans and perform joint work to deepen internal commitments and promote stronger internal engagement with practices and strategies meant to help schools adapt and respond to external demands for improved practice (Honig & Ikemoto, 2008).

Part of brokers' work may involve helping principals and teachers translate new demands (Spillane, 2000). Although, the presence of brokers does not ensure clear understanding given that demands may be vague or untested. Past research has demonstrated that when support is required by policy, but not clearly specified, as is the case with many interventions for low-performing schools, the support was found to be "limited and haphazard," and as a result, had little influence on teaching and learning (Finnigan, Bitter, and O'Day, 2009). By contrast, long-term improvements in school performance were sustained when support was high intensity (Supovitz & Turner, 2010) and tailored to a school's particular conditions, context, and needs (Marsh, Strunk, and Bush, 2012).

Past research has shown that brokers are not always successful; rather, their 'success' is often dependent on those whom they are supporting and hinges on others' engagement (Honig and Datnow, 2008). Brokers, too, are subjected to the challenges that exist in the complex and difficulties that stymie efforts to improve instruction and school improvement (Finnigan and Daly, 2012). Smylie & Corcoran (2006) suggest that the variability in the kinds of supports that brokers provide can be attributed to those they work with, the environment, or in some cases their own internal instability. Research has also shown that when brokers try to provide sustained support over time, their efforts may bump up against history and other improvement activities. When frustrated, brokers may revert to top-down tactics in their relationships with schools that run counter to the espoused flexibility and potentially dynamic capabilities that brokers might offer (Honig, 2006).

Top-down tactics rarely instill the kind of professional confidence that is necessary under a vendor-like arrangement where schools seek partners and service

providers – not more formal command-control. Successful brokers thoroughly diagnose problems and match supports to meet schools' core needs over time (Finnigan, Bitter, and O'Day, 2012). To match supports brokers may need to learn from their support efforts and make midcourse corrections (Ikemoto & Honig, 2008) or adapt strategies to fit local contexts. Brokers' adaptability further legitimates their professional authority with schools and where there is high degree of confidence in brokers' professional expertise, schools are more likely to engage with brokers' recommendations (Finnigan, Bitter, and O'Day, 2012). Although, the converse is true, that in instances where brokers are unable to adapt or match schools' particular needs, their legitimacy is in question and "school staff were likely to resist, adapt, or simply not implement the programs or recommendations the providers offered" (Finnigan, Bitter, and O'Day, 2012, pg. 16).

Research teaches little about what brokers do, or the role they could play in helping schools develop continuous improvement practices under Portfolio Management-like conditions. This is in part because support has often been selective, or narrowly focused on goals and strategies dictated by a central office. For example, in District 2 in NYC during the 1990s, administrators developed an instructional improvement strategy to be used districtwide. This involved coordinating learning standards, human capital, budget and assessments. Administrators in District 2 set explicit goals to measure progress towards instructional improvement; they created professional networks designed to foster a strong normative environment based on what the district endorsed as 'good practice.' Towards these ends, central office provided support for teachers and principals around decision-making under the new expectations for practice (Elmore & Burney, 1997).

While District 2 in NYC is held up as an example of a coherent system, it was created by central office. Central office took responsibility for organizing professional learning and coordinating that with key operational functions aligned to a principled framework. Coherence was expressed as standardized commitment to central office's continuous instructional improvement strategy, and to these ends, support was targeted to help schools align their internal conditions to fit with specified norms for practice.

We could expect that brokers would operate very differently under a portfolio management model given that central office is no longer coordinating instructional or

professional development; rather, principals are given autonomy to develop instructional systems and held accountable for the results. Under a portfolio management model, the expectation is that each principal is responsible for solving the problems that arise during efforts to coordinate instructional systems with organizational conditions. In turn, principals must figure out what professional learning opportunities may be needed for teachers to work in the school's new systems. As the key problem-solver, the principal must prove that she can hack it. Yet, under such conditions, it is harder to view accountability as formative, or a learning opportunity. This presented a key challenge for Networks, which were held accountable for their schools learning to improve, to help their schools see the accountability tools as formative learning experiences instead of just compliance tasks that could be completed in superficial ways. The challenge for networks was to help schools embrace accountability as an authentic learning opportunity – something rarely achieved under the result-oriented climates promoted by strict test-based accountability models.

Multiple accountabilities

The most prevalent type of school accountability is bureaucratic, test-based accountability that emphasizes students' test scores as measures of adult and system performance and treats the school as the basic unit of accountability (O'Day, 2002; Ladd, 1996). Although, education researchers have highlighted several forms of accountability that may operate within schools including professional, political, moral, and market accountabilities (Darling-Hammond & Ascher, 1991; Abelmann, and Furhman, 1996; Adams and Kirst, 1999; O'Day, 2002; Stetcher, Hamilton, and Gonzalez, 2003; Firestone and Gonzalez, 2013). These accountabilities may actually present competing demands when the values associated with each conflict (Finn, 2002). For example, one leader's sense of moral obligations to provide a safe, caring environment for his students may run counter to the testing culture that emerges from bureaucratic pressures.

Scholars have suggested that principals, in particular, enact accountability pressures (Spillane, Diamond, Burch, et al, 2002) as they interpret district mandates in relation to their personal leadership goals; they use their interpretations to define agendas and enlist teachers. "That is, they attend to some kinds of information and downplay

others, framing problems in particular ways. A principal's enactment of accountability might weigh personal, professional or political influences more heavily than bureaucratic requirements or market incentives" (Shipps, 2011, pg. 3).

For the most part, research has focused more on principals' enactments of multiple accountabilities than teachers, or support providers. In looking at a small sample of schools identified for sanctions in California, Mintrop and Trujillo (2007) found principals' orientations towards bureaucratic accountability played a significant role in their schools' organizational effectiveness. In higher growth schools, principals embraced the standardized tests and focused the school's internal efforts on helping students get higher scores on tests; in the bottom schools, external accountability was rejected and there were less effective organizational conditions (Mintrop and Trujillo, 2007). According to these authors, practical relevance of the accountability system for school improvement tended to occur in schools where leaders were open to external accountability and intentional about aligning commitments to specified targets.

Other research has considered how principals experience accountability and sought to understand what influenced their decision-making. Firestone and Gonzalez (2013) explored how middle school principals in New Jersey responded to NCLB pressure in relation to what may be considered competing interests like self-applied pressure to serve students, parents, teachers, superiors, other educators and the public. While the authors did not set out to test the relationship between principal's primary source of accountability and student achievement, they found that principals who led higher achieving schools "most typically had an internal sense of accountability" versus those principals in lower performing schools focused first on NCLB pressure and other external public pressure (Firestone and Gonzalez, 2013). The authors conclude that in higher performing schools (as measured by test scores), the principals were not necessarily immune to external pressures; rather, they "felt their personal code provided a way to reconcile cross-pressures of conflicting accountabilities" (pg. 399). This research seems to align with other research that argues that the social dynamics and collegial relationships operating inside schools matter more for improvement than technical responses to external pressure (see Finnigan and Daly, 2012).

Dorothy Shipps conducted a study of how principals in New York City experienced the reforms put in place by Mayor Bloomberg. While the sample size was fairly small (n=18), the findings suggest that several initiatives interacted in complex ways "that were not predicted by the policy-makers' theoretical assumptions" (pg. 3). It is also important to note that this study was conducted in 2007-08 – at an early stage in Mayor Bloomberg's reforms.   What she found was that few principals actually felt empowered; rather, more than two-thirds felt "beleaguered" by the mix of market accountability that came from the small-schools movement and the pressure to get high marks on the Progress Reports and Quality Reviews.

While all of the principals in Shipps' study reported some tensions, conflict was most acute among early career principals leading low-achieving, small high schools. Choosing a network support provider was not easy for these principals; they struggled more than principals leading large, politically embedded neighborhood schools.   Among the minority that reported feeling empowered, the principals either had a strong sense of professional judgment, which may have been bolstered by past experience or political clout, or they construed their decisions as being aligned to multiple sources of accountability, or legitimate demands from their support providers. What Shipps concluded was that NYC's combination of accountabilities could live up to its espoused merits, if more was "done to provide most, if not all, high school principals with the attitudes and skills currently helped by a minority" (pg. 33).

To date, there is little research that considers how support providers enact multiple accountabilities. Finnigan, Bitter, and O'Day (2012) argued that a market for support providers did not provide strong enough quality control among service providers, and as a result, the low performing schools in their study experienced limited benefits from external assistance. These authors suggest that there is little oversight of support due to limited capacity at the district and State levels. The burden of complying with the federal mandate to provide support for low performing schools often falls to districts more than State Departments of Education, but there are few restrictions on selecting providers, or evaluating their performance. In the absence of regulations, districts create markets for consultants and service providers.

The way that markets regulate quality demands free flowing, reliable information that 'consumers' can access and trust to use in their decision-making. In the absence of good information, or skilled decision-making, 'shoppers' may make ill-informed selections and the market does not regulate against low quality; inequitable service goes unchecked (Shipps, 2011). Bureaucratic accountability, by contrast, uses mandates to set targets and standardize evaluation; it uses monitoring and compliance to control for quality. Yet, we know very little about how support providers respond when there is a market for their services combined with bureaucratic accountability that holds them accountable for how well their schools perform as was the case in New York City.

In summary, research teaches us that schools with strong internal accountability systems are better positioned to respond to external pressures. Schools with weak technical cultures that privilege isolationism and preservation over collective learning are less able to respond to demands for improved performances. While brokers are seen as dynamic support able to target supports to help change practices in schools, they work best when the demands are clear and explicit. Brokers help to craft coherence between school's internal goals and the external demands, but their success depends upon the engagement of their school-partners. What the research is less clear about is how brokers are able to help schools see accountability as a formative learning opportunity and whether the combination of market and bureaucratic accountability leads to stronger supports.

Methods

The concepts described above grounded my analysis of the ways in which Networks used the Progress Reports and Quality Reviews in New York City. The Progress Report was used annually to evaluate schools by synthesizing multiple quantitative measures of student outcomes and school environment; it emphasizes student progress and compares each school to schools that serve similar populations of students. The Quality Review was a subjective, qualitative assessment of schools' practices, systems and structures and the extent to which they were aligned to strengthen teaching and learning. These two accountability tools were used by the Department of Education to make decisions about school closings, principal tenure, and to measure impact of network support. The espoused theory was that the tools would enable the Department and Networks to target

support, intervene in schools that need the most help, and make informed policy decisions.

This specific study uses an embedded case study design to explore how networks used these two tools in their work supporting schools. Drawing on the image of the networks as brokers permits me to consider the variable ways that networks targeted support in their relationships with schools, and where possible, to see how the accountability tools influenced those efforts. The goal with this study is not to provide generalizable claims; rather, the point is to explore how things worked in New York City and offer a starting point for future studies of accountability in action.

Data Collection

This study is based on data collected between June 2013 and July 2014. To understand better the work of networks in NYC, I shadowed one of the Cluster Leaders, which gave me access to the Cluster team and eleven Networks. To focus my study, I did an in-depth review of one network team's efforts to support thirty-three schools from thirteen districts spanning across all five boroughs. In this essay, I will refer to this Network as the pseudonym CFN 706. I selected CFN 706 because I was intrigued by the network leader's intense commitment to provide the best support for a diverse set of school partners in her network, which was evident from her contributions to Network Team Leader meetings that I attended. I was also encouraged by the Cluster Leader to work this network leader as she was seen as "one of the best instructional leaders in the Cluster."

The Cluster I observed saw themselves as a family that took care of each other. Groups select different ways to brand themselves. It seemed very fitting that this Cluster referred to themselves a family because, as a group, they seemed to possess a strong, collective ethical commitment towards supporting schools in *solving* problems. Within the Cluster there diverse perspectives and strengths, but they seemed to really value each other and their norms, which included attending to structures and protocols that helped them work as a team so they could, in their view, help schools given the ambitious aims and pace of reforms in NYC. This was managed differently in other Clusters, where formal titles and roles dictated interactions. Unlike central office, all of the Network

leaders in this Cluster "grew up" in the DOE and served as principals, superintendents, or regional coaches before they became Network Leaders. The Cluster Leader took great care to know the talent, disposition, and potential of the people who worked in the Cluster, networks, and schools. Seeing the Cluster as a family was important to the leader because he wanted people to be motivated to do what he often referred to as meaningful work – to "push thinking" and "move people" to deliver the best education to students – no matter what challenges stood in the way.

Certain network leaders in this Cluster were viewed as great operational leaders, some were more academic, and some were instructional leaders. CFN 706 was lead by a very strong instructional leader who had served as a principal, coach, and deputy network leader. She was a fierce leader, who held very high expectations for her staff and schools. She tremendous value in creating conditions that would promote strong instruction across all classrooms. This network was constantly working on informal and formal ways to strengthen instruction in-classrooms, which included building relationships with principals and school administrators. As one of the team members from CFN 706 expressed, "Without those relationships, you're not able to make any impact." CFN 706 wanted to make an impact and took the stance that schools needed strong technical cultures in order to improve.

My in-depth review of CFN 706 is an embedded case-study design (Yin, 2009) that includes data collected from the Cluster, network, superintendents, and schools within the network. Data for this research come from documents, records, audio-recorded semi-structured interviews with Cluster team members, Network team members, superintendents and principals, and notes recorded from several hours of observations of district, team, principal, and teacher meetings held in various settings across NYC; general impressions also come from the umbrella project that makes up my dissertation, which included sixty-eight interviews, approximately ninety hours of observation, and a substantial collection of documents that includes both internal DOE materials and thirty external reports about the reforms in NYC.

As I collected data, I pursued information that would provide me with a better sense of how CFN 706 organized its work as compared to other Networks within the

Cluster and offer clues to how the Network was mediating directives from central in its work with schools.

I believe it is important to note that I collected the data for this study during a particularly interesting time for Networks in NYC. As I mentioned elsewhere in this dissertation, Mayor Bloomberg's tenure came to an end in December 2013. Bill deBlasio won the mayoral election in November 2013 and was sworn into office in January 2014 - midway through my data collection.

This was an uncomfortable time for many of the participants in the study as it was unclear who the new Chancellor would be or what direction the DOE would take with regards to accountability and school support structure. As accountability and school support were clearly attributed to Mayor Bloomberg and his strategy of reform, it was likely these would change significantly. During his campaign in the fall, Mayor deBlasio promised a new direction, but it was not clear until months later what that might be. It was not till the end of April 2014 that the administration announced that networks would remain in place for one more year.

Newly elected, Mayor deBlasio issued a mandate to stop school closure, which took the teeth out of the Portfolio Management model. Amidst the unsettling, politically motivated changes, I found that people were still willing to share with me, but the tenor of the discussions and meetings were a little more tense than they were under Mayor Bloomberg's charge.

Data Analysis

I used Evernote to organize field notes and Dedoose qualitative software to code data in several phases. To begin, I narrowed my data to a subset drawn from the larger study that included nineteen interviews from Cluster and Network teams, superintendents, and principals. I sorted data into basic categories of actions performed by Networks generally including descriptions of various types of support they provided to schools, ways they organized themselves and how they managed relationships with central office and the Cluster. I also assigned a general code to instances where participants described ways Networks used the Progress Reports or Quality Reviews since the beginning of Networks in 2010. During the second phase, I recoded data to highlight when concepts I described

above appeared in the descriptions of activities performed by Networks including examples of when they acted as brokers negotiating engagement with the Progress Reports and Quality Reviews and when resources for learning and coherence emerged from their relationships with schools. I also catalogued data as "use" instances that captured important dimensions of how the Networks used the Progress Reports and Quality Reviews that did not obviously fit these categories.

Whenever possible, I relied on multiple sources of data to guard against threats to validity (Maxwell, 2005), which included drawing on external reports written about the Networks; I also relied on informal member checks with several participants to confirm my emerging sense of Networks and the varied ways they aim to support schools (Lincoln & Guba, 1985). Throughout the course of the study, I wrote descriptive and analytical memos to focus my inquiry towards answering my research questions (Miles & Huberman, 1984).

Findings

In this section, I discuss my findings related to the types of supports that networks provided to schools as they sought to help them respond to central office's demands to get better marks on the Progress Reports and Quality Reviews. I also describe the ways in which the networks, working as brokers with their diverse clients, used the Progress Reports and Quality Reviews in their efforts to support schools. I describe what I saw to be key factors that influenced the network team members' use of the Progress Reports and Quality Reviews including the impact of multiple accountabilities. It is important to note that these findings should be considered suggestive as I explored a sub-set of networks within one of NYC's five Clusters.

Types of Support

Networks provided numerous types of support that varied according to context, need, and demand. Some supports were urgent - fire-fighting in nature and were meant to help schools get out of an emergency situation. Others were focused on developing long-term instructional infrastructure with a goal of changing school wide culture. In some instances, support was requested by schools, and in others, networks planned for and provided supports that they believed were necessary for school improvement. In the

sections below, I highlight four common types of support that networks offered to their partner schools. The four common types include turnkey, functional, network-wide instructional support, and site-based instructional support.

Turnkey. Typically used as an adjective, *turnkey* is defined as: "complete and ready to be used." In NYC, the term was used as an adjective and verb. Because networks were seen as those who worked closest, and knew their schools best, the teams were seen as the repository through which all information was to reach schools. This system for information management was used equally by central office, Cluster team members, and schools. When central or clusters needed information to get to principals, it went through networks, and similarly, when principals needed information, they turned to their network first. Members of network teams were either assigned, or volunteered to serve as "points" for particular functional areas including budget, data management, teacher evaluation, English Language Learners, special education, and others.

Network point members were responsible for attending meetings hosted by either the cluster or central and then passing that information along. For example, one cluster leader explained how he worked with the points on the network teams in his cluster around issues related to the Progress Reports, "and so really, the role of the cluster, and my role in particular, is to work with the 11 different accountability points so that they could go in and directly work with their 18-32 schools. My job is to build their capacity to the point where they're capable of going in and turnkeying that information to the schools…it's literally impossible for me to visit every school and discuss the progress report."

A former network leader explained the role of turnkey: "Well, we always had different workshops and trainings – both as network leaders and even as principals. So, as a principal, it was given to use by the network. As a network, we'd get it from central or whoever was in charge of accountability. They'd come in with slides and go through the whole thing. And then on the network level, we always had one expert who went to all of the trainings…and they would go into the schools and work with principals helping them interpret data, or work with teachers – cause everyone had a data point in the school

who was supposed to turnkey the information to other teachers or workshops with parents."

Functional. As partners with their schools, network team members often served as an extra set of hands to help schools function. For example, in CFN 706 two of the team members rotated as a substitute principal for a K-8 school located in Queens because the school's principal was out sick and the assistant principal was not up to the task. For three months, the team members spent full days in the school to cover and help with teacher observations.

The district was in its first year of implementing a new teacher evaluation system (ADVANCE) that required multiple assessments of teacher practice. Every teacher received multiple observations based on Danielson's Framework for Teaching that had to be completed and uploaded into the Advance system by the beginning of June. In April and May, across networks and schools, team members were helping with teacher observations. One network leader referred to this time of year as "triage mode" as her network was really trying to help principals who "did not plan well, or execute well over the year to get everything into the system," explained the leader. Network 'support,' in some cases, involved team members doing some of the heavy lifting that schools needed done to remain in compliance - especially around times in the year when schools faced deadlines (like budgets which were due May 30th). Networks had to balance their long-term plans for support with cyclical cram-sessions when schools' needs were urgent matters; times of "triage" were particularly challenging as a majority of the networks' schools would be facing the same deadlines.

Network team members also provided schools with operational supports including help with grants, budgets, facilities, human resources, and safety. For example, CFN 706 helped one of its struggling middle schools from the Bronx apply for a School Improvement Grant. The school received the grant and was able to form community partnerships with Out-of-School Time Young Athletes, 21st Century Leadership Program, Oasis Summer Quest, and Bronx Opera to create more enrichment opportunities for the school's disadvantaged population of students (92% qualify for free or reduced lunch; 28% classified as SIFE). In addition, with the grant, the school was able to obtain

professional development opportunities with external providers like Facing History and Ourselves, Ramapo, Math Solutions, and the New York Historical Society; the network helped broker these relationships and assisted the coaches that were provided by through the partnerships. Another network was able to help one of its schools that had $278,000 in rollover debt reduce the debt by "cleaning up date and policy decisions about paying back the debt. The school got out of debt within the year," explained the network leader.

Other operational issues that networks provided support for included the day a middle school student was stabbed by another student outside of their Bronx middle school. Operational support was provided when there were union grievances and trials to resolve labor disputes. Operational support was provided to answer urgent calls when a school secretary entered measures of student learning (MOSL) or data for STARS into the 'system' incorrectly. Cross-functional network teams were meant to offer broad supports to help schools with any fire – no matter the size or intensity that happen to burn on any given day in a school across New York City.

Network-wide instructional professional development. Networks hosted a series of network-wide professional development opportunities. These opportunities were not mandatory functions that schools *had* to attend; rather, the networks designed these professional opportunities and encouraged principals to attend, or asked principals to designate teachers, assistant principals, or coaches to attend depending on the topic covered during the meetings. Between September 2013 and June 2014, CFN 706 hosted more than ninety professional development opportunities. Among these were monthly meetings for ELA and math teachers to work with the Common Core Learning Standards, as well as, sessions targeted towards the district's recommended curriculum (i.e., Go Math and Expeditionary Learning). Networks hosted principals' retreats each month. CFN 706 is branded as a strong network for *literacy* for early grades and middle school so there were also special sessions that covered "text complexity" in specific grade levels (i.e., 4-6 or K-2). Other networks hosted teacher job fairs to help their schools hire qualified teachers; there were various efforts by networks to develop network-wide support.

Network-wide professional development sessions served several purposes. One purpose was for the network to offer general help. A second was that sessions were designed to promote idea sharing and collaboration among the schools within the network. For example, CFN 706 hosted nine separate "study group" sessions for schools that addressed topics selected by the partner schools. Sessions covered a range of areas including "Nurturing Successful Kids," and "Engaging Boys Within Writing," to "Actionable Feedback for Observations and Artifacts" and "Using Brain Focused Learning Strategies." These study sessions were facilitated by network team members and designed to be interactive; the goal was to promote social learning among the participants. In addition, the network CFN 706 hosted more traditional professional developments, as well, with the goal of providing new ideas and enrichment. During the summer, CFN 706 arranged a three-day intensive seminar with Richard Elmore. Other networks provided similar opportunities for their principals.

Another purpose that is harder to capture was that networks sought to create a social support system for the schools in their network. For example, I observed the end-of-the-year principals' retreat held at Wave Hill, a famous public garden and cultural center in Hudson Hill – a neighborhood in Riverdale, Bronx (Teddy Roosevelt once lived at Wave Hill, as did Mark Twain). There were thirty-three principals and their guests sitting among cluster and network team members - all around grand, beautifully decorated tables in Armor Hall, a wondrous wing of the historic mansion; it felt more like a wedding reception than professional development.

At the retreat, principals listened to several presentations about RTi (Response to Intervention), a popular, early intervention strategy for helping students who are at risk of school failure. Principals were given an opportunity to engage with vendors, and learn about recent changes to the teacher contract, as well as, participate in exercises with protocols developed by the network to help principals take an inventory of the year that was coming to an end. These inventories were to inform principals' summer planning and help them develop strategic plans for the following year; this information was to be used by the network in developing targeted plans of support for each school. However, what seemed to be the most valuable parts of the day were the friendly conversations during lunch when people walked through the gardens and talked. It was an end-of-the-

year PD, at a time when principals were stressed about test-data being released, budgets, and the daily demands that they left unattended while they sat for a day of PD, but people seemed genuinely happy to be among colleagues; I think these observations suggested some of what made network-wide professional development beneficial in this network, where there were strong commitments to one another's success. These informal aspects were partly what drew principals to particular networks, which created an incentive for networks to find useful and relevant ways to cultivate informal networks and space in busy schedules for relationship building.

Site-based professional development. Site-based professional development provided by the networks emerged from planning sessions when network leaders would meet with principals – often at the beginning of the school year. During these sessions, the network leader, with the principal would map out what support might look like throughout the year. For example, one former network leader explained her early meetings with principals,

> One of the things that I did in my network, and I don't know if a lot of networks do this or not, when the progress reports came out, we made appointments at all of our schools as an instructional team with our data person and we really sat with the principal and pulled the progress report apart with the principal. Some principals pulled it apart for us. We didn't even really have to. But some principals really needed to have the team sit and look and say, this is what the progress report is telling you, let's figure out what our work here is going to look like for the rest of this year, cause they usually come out early November, so it gives us enough time to say, how are we going to revise our plan for what we're doing here this year?

The networks would map out this support for each of their and allocate to those plans a set of resources like coaches and time to make sure that plans were also executed. The task of managing site-based support plans fell to the networks. Initial commitments would shift throughout the year as other demands and urgent situations arose. The networks had to have organizational systems to track and monitor commitments. They often used a "support log" to track what was planned as compared to what was

accomplished. Team members used email to memorialize understandings and monitor progress towards plans.

Site-based support could take several forms, although, the most common involved coaches working with teacher teams, or assistant principals. An example of the way support could vary across schools is the way CFN 706 provided support to three schools that had identified "planning and implementation of standards-based Common Core aligned instruction within ELA and mathematics" as a school-wide instructional focus for the year. All three schools had the same focus, but CFN 706 coaches provided different supports based on each school's preferences. One of the ELA coaches from the network worked with one school in the Bronx on writing in literacy with a specific focus on grades 2-5. Over a six-week cycle, the coach meet with horizontal grade-level teacher teams and used Understanding by Design (Wiggins & McTigue) to focus on writing unit and daily lesson plans. At another school in the Bronx, the math coach from the network team worked with vertical-grade teams to conduct a Math Talk Study Group in which the teachers used lab-site and intervisitations (visit each other's classes) to align math instruction with the Common Core Learning Standards. Both ELA and math coaches worked at the third school in Manhattan, where the focus was on planning for Common Core aligned instruction in both subjects. Yet, even with these targeted, site-based kinds of supports, there were difficulties. Coaches reported that teachers in the schools did not continue to implement new strategies in-between coaches' visits, which limited the depth and range of instructional development among teachers.

In several interviews, network team members described how challenging it could be for them to achieve their network's goals given that they had no authority to demand anything of principals. These situations created an unproductive pressure on the network because their status and rating was on the line when their schools did not improve. It was unproductive because it was just pressure with no recourse to alter the outcomes, if schools would not cooperate with them. The networks could try to get help with their relationships with schools by reaching out to the Cluster to intervene, but similarly, the Cluster had no direct authority over schools. A network could reach out to a school's superintendent, who was the principal's rating officer, but a network's relationship with a superintendent could be tenuous given the oppositional transition under the new structure.

The networks' hands could be tied, if a school would not engage with the network to improve the school. One network team member shared, "You have those schools that just don't care and you can offer your support and offer your support and offer your support, and as a network system, we can't force ourselves into the schools. If a school doesn't invite you in, what do you do? You can send 20-30 emails, like I said, a monthly email, every three weeks, really, saying here are all of my open dates for the next couple weeks, let me know when you want me to come in. And the same seven or eight schools don't respond - half of them TAP schools (identified as priority or focus schools for low performance)."

One team member from CFN 706 echoed this sentiment by describing how that network invites people to attend network-wide professional development sessions and they reach out specifically to individual schools to provide on-site assistance. For example, he explained, "We have one school in Harlem - that principal did not send anybody to anything. I sent a good 80 emails over the last 6-10 weeks saying, 'I have to come in. I want to meet with some of your teacher teams. You're a TAP school. You're an IA (interim-acting principal), you need to show that you're giving support.' And she always had an excuse for why. Finally, I got in there yesterday. Sometimes, it comes down to Kristin or Richard (leaders) - a lot of times it's Richard because he's more of a gentle speaker than some other people on the team, but even then, it's not always successful. But that's the nature of the system. If they want to pay $40,000 just to get operations support and a couple emails with attachments, they can do it. We can keep trying and trying and trying, but it's their choice ultimately."

One former network leader explained, "as a network leader, you are not a supervisor. You cannot tell a principal what to do or how to do it. You have to really try your best to develop a good strong relationship with the principal and really try to support their work, but there have been occasions where I have suggested certain curriculum that they should use or programs or scheduling or so on and when I've gone back to see that those things were implemented, they weren't. So, there was no follow-up and that happens on occasion."

She went on and explained that whether support is successful can often depend on "…the strength of the principal and depending on the strength of the structures that they

have in place - their instructional leaders or coaches in the school. If they haven't built capacity, which is what we try to do, I mean we go in and try to work with either an AP or a coach or someone to really be to - so that when we leave, they can continue this, but the principal has to be on-board and they have to ensure that they're putting the structures in place for this to happen, and that's when I become frustrated sometimes, as a network leader." Working in close relationships with schools could be a source of significant stress – especially in struggling schools, where there was a lot of heavy lifting to be done. Under the empowerment structure, principals had autonomy to set the vision for the school and create strategies for improvement. Despite the new systems in the district, there were still some principals who tried to operate in old ways like buffering against outside influence. Unfortunately, this also meant that they sought to buffer their schools from the networks, too.

Networks' use of the Progress Reports and Quality Reviews

From interviews and observations, it was evident that the Networks relied on the Progress Reports and Quality Reviews to orient their work internally and in their relationships with schools. In the section below, I highlight how networks used schools' outcomes, or grades, on the Progress Reports and Quality Reviews, to prioritize and organize how they would attend to the task of providing support; networks discriminately made strategic plans based on the schools' scores to direct attention and allocate resources.

Organizing in this way created work, and sometimes tension, for the networks as they were supposed to provide support upon principals' requests. In some cases, their priority support plans conflicted with principals' wishes – in particular, this kind of tension arose when principals who were serving in high-performing schools wanted to maximize network support. There were times that the networks had to stall high-performing principals, who may have been requesting support, but did not rank has a high-priority in the network's organizational scheme.

This kind of tension created a double-jeopardy for networks. In a market environment, delivering untimely, or less than desired support to high-performing principals was a risky move for networks. If a network failed to satisfy a high-performing principal and the principal chose to leave the network, the network's NPM

score would suffer as the aggregate of schools' marks was weighed heavily in the NPM. Losing a high-performing school was less than ideal, but the low-performing schools needed significant support that demanded more resources than were available. Managing this tension was more important for some networks; the intensity was determined largely by the configuration of schools in a network.

Another set of key findings involved the ways in which the networks used the Progress Reports and Quality Reviews in their relationships with schools. A common practice was to use the tools to develop shared plans of support in planning with principals. However, the use of the tools also varied across schools, networks, and tools; there were some networks that helped their schools get a quick 'bump' on the Progress Reports and others that resisted the low-hanging fruit of the Progress Report to focus on more long-term supports directed at changing the instructional systems and organizational conditions in schools largely based on the Quality Review rubric. I argue that these different practices represented different approaches towards support; some types of support were meant to address technical, where as others were meant to address adaptive challenges in changing school performance.

In some ways, this represented an ironic dichotomy that was built into accountability metrics. The metrics for the Progress Reports were proxies of performance meant to signal the level of quality of educational experiences students were having in schools. Yet, these measured students' performances. By contrast, the Quality Review measured more precisely actual, day-to-day performances of adults in the schools. Schools, with networks' help, could address low performance on the Progress Reports without really addressing the adults' performances in the building. Focusing on the technical challenge (low-hanging fruit) presented by low scores on certain parts of the Progress Report seems like an ironic form of 'support' for school improvement, but it did happen, and was incentivized by the designs. In the following section, I present more evidence of these findings.

Using Progress Reports and Quality Reviews to organize networks' work

There were no directives from the DOE outlining how networks should organize themselves; rather, the networks had autonomy, in some ways, to determine how they

would deliver support and organize themselves. The Cluster team provided some help with sorting this out, but network leaders, especially, if they were successful, were allowed to decide how to allocate coaches and manage how they would deliver supports to principals. This created opportunities; networks could be flexible and responsive. However, the autonomy created challenges because the networks had to stay on top of everything. One network leader explained, "You gotta keep up with stuff because nobody tells you. Sometimes you've got to learn and find out on your own."

To help networks, central office created the Children First Intensives (CFI) to turnkey information to network leaders and provide some general support for the teams. These monthly meetings were treated as professional development sessions for networks. Some portion of each meeting usually involved a sharing of best practices among networks. During the fall of 2013, October's session was titled, "A culture for learning." The session explored assessment landscapes and "strong culture for learning." According to the DOE staff member who introduced the session, it was intended to "move way from simply information-giving towards professional development." As she explained the scope and sequence for the rest of the year, she shared: "The Network Leader job is very difficult because there is a deluge of mandates and expectations; central realized they had to figure out how to organize and prioritize, but still keep their eye on the target. The goal [of CFI] is to focus people instead of getting distracted by the noise in system." It was up to the networks to find ways to stay focused; they often turned to the Progress Reports and Quality Reviews to focus attention and prioritize their work.

In interviews and observations, it was evident that networks created tracking systems to organize their internal operations and plan, as a team, how they would provide support to the diverse schools in the network. There were several factors that influenced how a network organized its work, including the network's assessments of its principals, schools, and the quality of coaches on a team.

A common approach that networks used to organize themselves was to create tracking charts, but their charts looked different, and networks used them differently.

On a day in November, I traveled to a network office located in the Bronx across the street from Fordham University. In the office, hanging on the walls were white butcher paper that contained lists of numbers and colored post-it notes. The numbers

looked like a military code; really these represented the way NYC names its schools: 08X093, 07X031, 09X274 and so on. The first number is the old district number (telling you location), the letter signals the borough (i.e., X stands for Bronx; M stands for Manhattan) and the last number is the school number.   Schools also have names like Stuyvesant High School, or Central Park East; insiders know when to refer to a school by number or name. In the network office, on the paper on the wall next to each number was a series of post-it notes that was either green, yellow, red, or purple. The network leader explained the system:

> What we do in August or as soon as we get, whatever data we have because some come in later than others, so then what we try to engage in is then we put all our schools down, we put their focus, whatever, cause this year the Department of Ed asked them to have an instructional focus and then we just kind of like, based on the data, based on the interactions last year, we try to sort of prioritize them, but we kind of just went through and said, these people will need intensive, these are the priority schools, these may have sort of a middle touch, the other ones are a check-in type thing, so we use a different way to, but we sort of had like where we're all sitting together and putting our two cents in.

Red post-it notes were used for "intense priority," yellow for "middle touch," and green for "check in." This network was providing support for thirty-four schools during the 2013-2014 – this was up from twenty-two the year before. Prioritizing which schools needed the most attention was extremely important for the network to keep track of resources, coaches, time, and plans.

Similar to other networks, this one adapted plans throughout the year as it received more information; post-it notes were rearranged. For example, the network leader explained, "we adjust constantly because like right now we have a batch of, that's what I'm coming in here to do today, was to go chart these QRs and then make some teams and now we have to shift, we have also some state reviews." The DOE announced annually in October which schools would receive a Quality Review or an alternative Quality Review (to be managed by the networks); networks adjusted accordingly.

CFN 706 managed its tiering systems differently. They had a data analyst on their team who created a tracking tool that they referred to as the school's 'portrait.'

Portraits were listed using an excel spreadsheet that contained information about each school's Progress Report grades and Quality Review rankings for the last three years, State Report Card data and AYP status, and TAP information (Targeted Action Plan). This network leader explained how her network uses the data and consultations with the schools to organize and tier its supports:

> So what we do is, in August when our state exam scores come back and the progress reports come out, we do an onsite consultation with each of our schools and depending on where they fall in that matrix, on the quality review, the progress report, in fact, Jerry*, our data person, creates a spreadsheet of every school. And on that spreadsheet are the quality review for the last three years, the progress report for the last three years, performance levels, progress levels, and then when we go for consultation, it lets me, as the network leader, and even when I was the deputy, decide who needs to sit at the table. Some of it is dependent on the demographics of the school. Some of it is dependent on the quantitative data from the state. And what we do during that consultation in August, we basically review all the data points and we create a prioritized action plan and assign a liaison to each part of the action plan to the school. Sometimes it's by content specialty, so if the school had a drop in math and maybe the math achievement specialist, sometimes it has to do with relationship, you know you know that the school has an outside math staff developer and doesn't really need the content specialist but maybe needs more leadership support for themselves or the assistant principal so it could be me or in the past it was Susie* or Richard* and we provide the leadership support, so during that consultation, we try to kind of flush out how the various data points work within the context of the school and the school leader, which is the part that all of those data points don't tell us, that's where that kind of emotional intelligence on the part of the network fits in.

This network used the Progress Reports and Quality Reviews to inform their assessments of schools' needs, but also has an entrée into discussions with schools. Principals were not required to consult with networks; the network arranged those consultations under the guise that it would help schools narrow their focus. The portraits provided networks a legitimate ice-breaker for conversations with schools. Networks had been told, as one

network leader explained, "if principals don't call you, you back off...so it was a little bit of, there's bravery, calling, and I did a lot of those calls."

In many ways, the timing of the Progress Reports and Quality Reviews dictated when networks would interact with their schools. Progress Report grades were released to the public in October. The networks received their schools' grades in advance of the Report Cards being released to the public. Once networks received the data, they would reconcile that information with what they knew about the schools and decided whether to contact their principals. One network leader explained, "Well, if you got a D and then being whatever I know from my team members, cause I have someone in every school and my own experience, so I might have a sense of how we can assist them [principal] or what I think they [the schools] need to put in place but it's just really more of the first initial conversation is just kind of preparing them [principals] to absorb the shock because what has happened in New York City because of the report card grades, you kind of own it, it's a very public thing and you take it like your grade as a principal cause even - as we are all former principals and we all felt the same way, that if a school got a C, it was like, 'oh my God,' I got a C. It's like you're getting a C in your class. So I just really try to message to them ahead of time so they're not shocked."

Quality Reviews were conducted between December and May. Often, networks helped schools prepare for Quality Reviews by conducting 'mock reviews.' One network leader explained, "as soon as we find out they're going to have one, it will almost be a weekly visit…like weekly touch-ins with schools." Alternative Quality Reviews were different because those were managed directly by the networks.

The networks used the Progress Reports and Quality Reviews to assess schools' needs, however, they rarely relied on overall marks; rather, they looked at individual components of the tools. Networks were well aware of how misleading a Progress Report grade might be. Because the Progress Report grades were based on comparisons of similar schools, an A might not reflect top-quality school – rather, it signals that school might be the best of a low performing group. For example, one network team member offered, "Quick example, in the South Bronx, we have a school that's a Triple A school, but this year had 10% of their students on grade level but they were the number two school in their district, every other school had two and three percent. And so when you

think about it, 10% sucks, but you're four or five times better than your local schools and that does say something, I think."

Given the design, networks resisted using only grades. A network leader explained, "It depends because it's a progress score so they still may have low state test results. They still may be on a state list with the A, so they still need support," but it could also be that a school was low on the Progress Reports, but "they may have systems to support the progress of kids, intervention system, RDIs, the effective use of their extended day, so they may have put things in place, they may not be as intensive as some of the priority schools that come up because of the D and F cause those TAP plans require time and energy so then you're really, you focus on that." The 'systems' that schools have in place are reflected in the Quality Review; it was important to look at both tools.

In the networks I observed, they diagnosed 'need' based on the information provided by the tools, as well as, their in-depth knowledge of their schools. For example, one network leader explained what she considered when determining which schools needed supports. She shared that she had some schools that had A's on the Progress Reports, "but you can't leave those people. It all depends on the leader cause we have like someone like Jane*, who's had A's, four, five in a row, but she has capacity in her building. She knows how to lead. So, if we need to pull somebody, they might be alright, they're not going to touch-in on her because she's got it under control, but then you may have another person that got an A, but they still need the other kind of supports."

The networks developed informal mechanisms for assessing their schools including in-person visits with schools. Some networks had every team member walk through all of their schools any where from two to three times a month; other networks had a majority of their members work from their cubicles. The social interactions seemed to matter for assessments of needs, as well as, for the depth of supports provided. When I asked one team member how his network assessed schools to tier support, he responded,

> Anybody, you can walk into a school and within an hour, you know if that
> school's well-functioning or not. There's just tell-tale signs. Is the principal only

in her office or his office? Are the APs anywhere to be seen? Are there kids running around? You walk into ten classrooms and is there consistently ten classrooms of students being instructed? I mean, there's just basic things. Do you see good student work that looks plagiarized? And, I think if you ask, I'm pretty certain actually, if you asked our team to rank order our schools from messy to least messy, you'd get the same top ten from every team member. I think ten to twenty might be a little bit off, in different orders, and then I think 20-30 would probably be consistent again, the most effective schools.

This team member expressed that what he believed would have been shared assessments was due to the fact that every member of the team visited most schools in the network at least twice a month. In addition, every Monday, this team met to discuss their schools. The knowledge that was gained through social interactions with during visits to schools, and the social learning that occurred among network team members during their weekly meetings, was used to develop and adjust plans for support.

Honig, Copland, et al. (2010) refer to this kind of individualized, need-based support as a "case management" approach to support. According to these authors, case management stands in contrast to traditional central office support because support "became experts in the specific needs, strengths, goals, and character of each individual school in their case load and worked to provide high-quality responsive services appropriate to their individual schools" (pg. 70). A case management approach and "problem-solving through project management" helped networks specialize in serving schools' needs instead of specializing in general services that may or may not help their partner school. For example, one network leader explained, "what I did was basically, we would look at the data and then we would decide which coaches would be assigned to which schools based on what the needs were, whether it be math or science, or ELLs, or someone."

Networks had autonomy to decide how they would allocate their coaches.   The decisions could be based on a variety of factors ranging from a particular coach's content knowledge, or interpersonal relationships with particular schools. For example, one former network leader explained, "I mean they [coaches] all have different - Ed was great with teacher teams. Sandy not only did math, but she is also great with the early

childhood grades…sometimes it was relationships that if we needed to send someone in cause they were having a hard time with something and they needed someone they had a good connection with - like Kristin (deputy network leader) - anytime Andrew (principal) might have had a problem, she was the person to go see him cause she could kind of get through to him."

The networks used the data and processes associated with the Progress Reports and Quality Reviews to assess schools' needs and make decisions about how the team should respond given certain constraints and considerations. These assessments were conducted through formal and informal mechanisms that varied across networks, and resulted in varied and targeted supports offered to schools.

Using the Progress Reports with schools to meet technical challenges
By design, the Progress Reports and Quality Reviews asked schools to account for different aspects of their performances; the difference in nature of these evaluations created distinct kinds of opportunities for support. Every school received a Progress Report every year; Quality Reviews varied and depended on a school's circumstances, including a school's Progress Report grades. For example, if a school received its third C in a row, it would get a Quality Review that year. Every school received a Quality Review at least once every four years.

Annual Progress Reports looked at schools' quantitative data like test scores and students' credit accumulation. By contrast, the Quality Review evaluated schools on the systems, structures, and processes they had in place to improve instruction. Addressing low 'performance' as evaluated by the different metrics on each tool demanded very different strategies for improving marks and different kinds of support.

While the Progress Reports were new in the sense that schools were judged on how well their students performed as compared to similar schools, the Progress Reports were not radical departures from prior evaluations, or State evaluations, that judged school performance by students' test scores. Given that the Progress Reports evaluations were familiar, responding to them was a technical challenge for most schools.

I consider these *technical* challenges based on Ronald Heifitz's definition of technical challenges as the kind of challenge where the solution resides in current ways of

174

working. This does not mean that they are easy challenges – just that they are problems with known solutions that can be implemented through current 'know-how' (Heifetz, Grashow, and Linsky, 2009). Technical problems "can be resolved through the application of authoritative expertise and through the organization's current structures, procedures, and ways of doing things" (pg. 19).

Several networks specialized in moving schools by helping principals attend to what was seen as 'low-hanging fruit' on the Progress Reports. I want to avoid suggesting that they 'gamed' the system because the networks that emphasized this improvement strategy were still guiding principals to focus attention on a school's lowest performing students, who in the past may have gone unnoticed. Instead, I want to highlight that this kind of support was one in which some networks developed expertise to help schools' improve Progress Report grades without really attending to the quality of instruction in classrooms, or the organizational conditions related to improving instruction.

One former network leader, who was working on the Cluster team, explained to me that as a network leader, he would help principals identify the things you can control on the Progress Reports and then help them do everything within their power to attend to those aspects. For example, "…As and Bs – that is the most important… The instructional work of the Quality Review, well that is longer term work and there's teacher turnover. If you get an A or B on the Progress Report, it doesn't matter what you get in the QR. It's not hard to get a B, if you figure out which pieces are within your control."

This network leader reported that he would advise his principals to attend to the "scholarship report" that gave information about which students were passing their classes. He explained, "that's the single biggest indicator – are they [students] accumulating 10 or more credits? They have to pass. Schools that do this well sit down with every teacher and talk…we can't let kids fail." He talked about moving students off his register, and encouraging students who were not going to graduate to move to a transfer school: "That impacts how many kids are in the denominator," he explained.

Networks acquired technical expertise by knowing the Progress Report methodology really well. Then networks would help principals see what was within the

principal's immediate control and short-term, quick fixes that would produce 'improved' scores on the Progress Report. For example, one former network leader explained:

> If you look at a progress report you'll see there's only a bar for social studies, well there's two exams in social studies, there's three exams in science, there's three or four in math. So what they do is, and this is where it's kind of unreliable in my mind, but what they do is they look at historical data trends for a given demographic and then they assign a weight to that particular student for each of the given exams and each exam the historical data is looked at separately. So a kid that comes in as a level one, that's black, that's free lunch, could have a weight of like nine. And then a kid that comes in at a level four that's Asian and not free lunch might only get a decile of one or two. And then based on the data they make predictions about how if the kids are going to pass and what they're going to pass with and then what the kid actually performs you're measured against that, the average generates that prediction. So it's not really progress in my mind. Again, this is just something that schools do and this is certainly something I did, is that, a lot of schools don't do it, but smart schools do do it. You can find out precisely what every kid's decile weight is in your school and so one of the things I did is, I would find that out.

He went on to explain that he would find out the kids' decile weights and then track kids closely. Tracking students closely allowed the networks and principals to know which students needed to pass which tests, and which classes students needed to take, or who needed to go to summer school.

This version of close monitoring and "educational triage" is slightly different than what Jennifer Booher-Jennings (2005) found happening for "bubble kids" in Texas schools. Booher-Jennings found that teachers and school leadership sought to create an appearance of school improvement, as would be evidenced with higher test scores, by diverting resources to "bubble kids" – those students on the threshold of passing the Texas Assessment of Knowledge and Skills test. What made this version different was that the networks were encouraging principals to attend to the lowest achieving kids, who in Booher-Jennings' study were ignored. The kind of tracking encouraged by some networks as a way to respond to the Progress Report required principals to seek new

information, but this form of problem-solving – as a response strategy – did not really push principals beyond the bounds of traditional school management, nor did it require vastly different beliefs or ways of working, or managing for instructional improvement across a school because they were able to make superficial changes that minimally affected the school's structures or culture. For example, a common solution was to send targeted students to after-school programs. In these instances, the 'solution' required no examination of practice to consider how it could be improved. Instead of using the information from accountability to inform continuous improvement practices across the school, these kinds of solutions were student-based and targeted specific students.

Another illustration of this kind of technical problem-solving came from a principal who explained how she achieved an A by dragging a kid to summer school because the student had 9.5 credits, or "if a student had 89.8 on average of Regents exams, because that will get you a different kind of diploma, I would just beg the kid to retake one regents," explained the principal. She also told a story about how she was once able to change a B to an A on her school's Progress Report: "In the middle of one night, I got up at three o'clock, I got a B, my old school. And I went in to look at the number how it says, how many out of 95 graduates, I noticed that those three students [who had gone to summer school] were not added in. So I emailed Jim Lieberman, three o'clock in the morning, said I can't believe after a million conversations, these three students were not included in this file. And he emailed back, you are right. So my grade changed from B to A." Other network leaders shared stories about moving schools from C to A, and there were reports from network leaders of moving schools by providing supports like curriculum and teacher teams. For example, one network leader explained, "We have, you know, schools that had a D that we moved through targeted plan and the supports, they moved to an A, a D to a B," but she acknowledged, "I think it's a numbers game if you could just focus on targeting those specific kids who are very low and ensuring that they learn more, pick up more skills." Seen as technical problems, the solutions provided for the low students often included more after-school tutoring or test-prep for Regents exams or state assessments in the lower grades.

The networks that emphasized technical solutions also discussed the trade-offs involved in that kind of support. For example, one former network leader explained, "I

don't know how much you want everybody to do this because sometimes you do lose insight on instruction." Yet, this network leader also reported that the help she provided to the principals in her network was "well, I just want to say, it was so well received." The Progress Reports were very public and easy to read – on purpose. Like Zagat's restaurant grades, an A on the Progress Report was simply symbolic, but it was also substantial. When schools' grades improved on the Progress Reports – even if it did not mean that the quality of the school improved, it made it easier to face the public and community with a better grade.

Using the Quality Reviews with schools to meet adaptive challenges
Networks, as autonomous support providers, were able to target support and decide how they would balance support based on Progress Reports, Quality Reviews, or some combination. As mentioned above, the Quality Reviews presented different kinds of challenges for schools, and in turn, for networks to provide adequate supports. I considered these *adaptive* challenges as the Quality Reviews required people inside, and those supporting schools, to attend to school improvement by looking at the coordination of systems to support instructional improvement. While some schools operate in ways aligned to such thinking, they had not been held accountable for such behaviors until the Quality Review was introduced. In many ways, this required different ways of working both inside schools, but also in school support relationships.

I draw on Heifitz's conception of adaptive challenges as those that "…can only be addressed through changes in people's priorities, beliefs, habits, and loyalties. Making progress requires going beyond any authoritative expertise to mobilize discovery, shedding certain entrenched ways, tolerating losses, and generating new capacity to thrive anew" (Heifetz, Grashow, and Linsky, 2009, pg. 19). In a review of district-based improvement reform efforts, Jonathan Supovitz (2006) acknowledged that the distinctions between the challenges can at times appear to be a point of framing, and often times one that is missed – people try to solve adaptive challenges with technical solutions all of the time. According to Supovitz, understanding the challenge is paramount to getting solutions right. He wrote, "helping minority students pass a test is a technical challenge; reducing the achievement gap is an adaptive challenge."

Networks relied on the Quality Review rubric as a resource for designing professional development opportunities for their partner schools. Often, they would focus on particular aspects of the rubric that they felt addressed their schools' areas of greatest weakness, or the most challenging aspects of the rubric. For example, one former network leader explained,

> I used to look at the Quality Review. I'd see that it would say, 'lack of differentiation or multiple entry points.' I would say, 'you know what we really need to do - a lot of professional development for teachers on how to differentiate for students - how to meet their needs and scaffold.' Then, I saw that it said, 'lack of rigor; there were not enough high-level questions.' So, we did a whole piece on DOK (Depth of Knoweldge). We would look at trends. There would be specific things that when we sat down with principals in August - things specific to that school, and then we would make that a priority. Also, we would look at trends and that would be the professional development for the entire network.

In this way, the Quality Review rubric, in particular, offered a framework that networks used to design supports for developing their schools' in-depth engagement with the Quality Review rubric.

One network within the Cluster championed the Quality Review rubric as *the* support it would provide. Each month, this network hosted 'principal conferences' during which the network worked through, "Your School, Your Story," which was a tool the network created based on the Quality Review rubric. The network developed the "Your School, Your Story" tool to focus principals on what the network designated as three essential components of the rubric: indicators 3.1 (goals and action plans), 1.1 (curriculum) and 1.2 (pedagogy). The other seven indicators were valued, but not considered as closely. A team member from this network explained her use of the tool with principals:

> We meet at a different school each month. I would fishbowl a conversation with the principal around 3.1, 1.1, and 1.2 and we'd spend 45 minutes where the principals got familiar with the kinds of questions the reviewer asks in those three areas and really give the principal a chance to tell, we kept referring to, the story of their school. Who they are, what they're about, what this looks like in terms of

a student's experience in the classroom in their school? Other principals would take notes on what the principal's saying and at the end of that conversation, before we went to classrooms, we'd clarify what we heard the principal say in terms of what are the things happening in classrooms that are distinct to your school.

There were several benefits schools in this network gained from this exercise including that these schools outperformed schools across the city on indicators related to the instructional core. For example, Indicator 1.1, which is considered one of the "instructional core" indicators, examines how well schools ensure that curricula in all subjects are engaging, rigorous and coherent, accessible to a variety of learners and aligned to Common Core Learning Standards. On this indicator, in 2012-2013, 85% of the schools in this network were proficient or better; whereas only 52% of the schools citywide were doing that well. Schools in the network also outperformed the city on indicator (2.2), which attempts to measure how well a school aligns its assessments to curricula and how well those in the school analyze information on student learning to adjust instructional decisions at team and classroom levels. 85% of the schools in the network earned proficient or better on 2.2, versus 54% across the city.

The coach from this network expressed that her network really intended to help principals develop a "QR mindset." Developing a QR mindset, or focusing on the Quality Review rubric was seen by networks and schools as taking the long view. A QR mindset was a longer-term strategy for school improvement, which required principals to understand how systems in the school were working together. For example, one Cluster member shared, "The instructional work of the Quality Review, well that is longer term work." The instructional work of the Quality Review was focused on curricula, pedagogy, and assessments; school culture; and structures for improvement – and the interactions among these three domains. When principals developed a QR mindset, they began to treat managing improvement like a Rube Goldberg machine – a systems approach the considered interdependencies. Similar to building Rube Goldberg machines, this kind of approach to improvement required patience and strategic planning, as well as, ongoing commitments. For example, in a discussion between the Cluster leader and a network leader, the network leader was describing how her network

supported the twenty-five schools in her network received a Quality Review in the 2013-2014 school year. The Cluster leader asked her, "How much of this was you supporting and how much was prepping?" She responded, "well, it's both to some extent because the rubric is so useful and the work is on-going. It requires the practices that we want the schools to continue" around teacher teams, data use, leadership. She explained, "it's beyond prep" because they want the work to remain ongoing and they continue to support the teacher teams well after the reviews. Yet, the pace of this kind of change and improvement is very different from marching three kids off to summer school.

Using the tools to develop with schools shared plans of support
Several network leaders talked about using the accountability tools in one-on-one meetings with principals to interpret the school's Progress Reports and Quality Reviews in order to build a shared understanding between the principal and network team. These meetings were positioned as inquiring into a principal's goals for his or her school based on the evidence provided by the Progress Reports and Quality Reviews. In addition, network leaders used these meetings to determine with principals what the terms of support provided by the network would comprise throughout the year.   For example, a Cluster leader explained, "Whatever the need is for that particular school, you need to be able to call it out. Meet with the principal in August and have that consultation. Talk with the principal about what their instructional plan is going forward and what the network support would be and who's the best match on the coach side?" Network leaders needed to be able to facilitate these conversations; they used the Progress Reports and Quality Reviews as evidence of "the need for that particular school" that legitimized their calling it out.

Interviews suggested that the networks worked jointly with principals to analyze evidence and articulate clear goals – for the school and the support relationship. For example, one network leader explained,

> What we try to do is every August in the summer we meet with our principals and
> their cabinet and we look at, what are the goals and the priorities for that year? So
> if we look at the progress report, we'll see that as far as progress the students are
> making progress so whatever interventions and whatever they're doing with Level

Ones going to Level Twos is definitely strong and we want to continue with that. We'll look at subgroups to see maybe if there are certain subgroups though that are not making adequate gains so we may think about what are we doing for our L populations? What are we doing for students with disabilities? And then talk about what the goals are: What are the programs? Who's teaching those classes? And looking at all that. And then we'll look at the performance and then we will target that and say, okay great, you're an A school, you have a great environment, the culture is good, you're moving your Ones to Twos but you're not getting those Threes and Fours so what are we doing here? What's the rigor? What can we do for our Twos to get them to Threes? We'll talk about what supports we can give them, what they're doing in the school, so we will always break down the entire report and not just say, you're an A school, so you're fine. So we have to look at the realities of it and that's the way we look at that. And then we'll bring in the quality review as well and look at what those indicators are telling us and I think I like to look at that to see what's in the classrooms and what's happening as far as teaching and learning.

I heard across interviews descriptions of collaborative efforts to develop shared plans for improvement collaboratively was echoed across interviews. Networks engaged principals in analyzing student data – as seen in the quote where twos, threes and fours refers to students' proficiency levels. This kind of hands-on diagnosis was done even with A schools, where perhaps networks could have backed-off since the school was getting top marks. Across schools and contexts, networks were pushing the principles of continuous improvement with their partner principals.

There are two other illustrative points from the quote above. The first is that this point further demonstrates how networks relied on both tools to determine supports for schools. Networks required their principals – even Triple A principals – to analyze the sub-components of the Progress Reports to target areas of need: "… so we will always break down the entire report and not just say, you're an A school, so you're fine. So we have to look at the realities of it and that's the way we look at that." Secondly, they looked to the Quality Reviews because they offered a better sense of the structures related to instruction.

The other important point evidenced by the quote above, which was expressed in several interviews, was that network leaders used these meetings with principals – bolstered by the analysis and evidence – to develop shared plans for support: "We'll talk about what supports we can give them, what they're doing in the school…" For example, one former network leader explained, "…through the work of the progress report, we were able to help principals collaborate - not help, collaborate with principals to identify focused areas for targeted improvement."

Network leaders, team members, and principals would develop "what our work here is going to look like for the rest of year" as explained by one Cluster team member. While plans would adjust throughout the year as networks adapted to schools' needs, adjustments were not made without consideration the Progress Reports or Quality Reviews.

Lastly, networks developed shared plans for support using the Progress Reports and Quality Reviews – as well as other data tools, but the process for developing those plans would be adjusted based on the principals' experience, knowledge, and disposition. In some instances, the networks were equal partners with their principals; in other instances, the networks served more of a coach role, and others more of an instructor. With some schools, networks had to help principals develop an inquiry approach to improvement. For example, one network leader described a time she was helping a principal review state test data. She described what she was thinking:

> Why don't you unpack that and see how your children did with some of those
> power type of standards? And then prioritize. Now, of those thirty-five, all of
> these numbers, to say what are the focus, what do you really need to work on
> across the school? So, flush it out so it's not just a page with numbers…and then,
> are there any trends? Is there anything we see that's school-wide? So, you have to
> kind of do a thoughtful analysis – not just paper with numbers. You know what
> questions to ask of the data. That's the kind of thing – you have to help them
> [principals] – like 'what is that telling you?'

In some cases, networks needed to help their principals "know what questions to ask of the data" in order to come up with a plan for improvement. Schools moved through different phases of developing plans at different paces; the networks needed to be able to

help principals – in whatever phase they were in. Sometimes, identifying a "particular need" required networks to distinguish leadership needs from school needs as there were times that leaders were ill equipped to develop plans, or possessed ill advised ideas. The networks had to be able to manage the nuances involved in the different situations, which could be very different across schools in a network.

While the Progress Reports and Quality Reviews were helpful to networks as they developed shared plans of support, the tools were limited for they could not help networks navigate the interpersonal, relational aspects that went into developing shared plans of support. Tools may be useful, but experiences are shaped by tacit knowledge and informal conditions like trust, which were created through engagement with the tools and social interactions.   It is for this reason that one of the Cluster team members argued that the Quality Reviews and Progress Reports should not be referred to as "tools." He explained,

> …we don't use the word, tool, correctly. So a hammer is a tool and if you hit the nail on the head, you're going to drive a nail into a piece of wood with a hammer. It doesn't follow in social science, it just doesn't follow. You can have successes and failures that are beyond your ability to explain them or at least beyond your ability to explain them based on the six hours a day you're with the people that you're with. So I know we use the word, tool, but a tool is a precise instrument that does a precise thing the same way every time. These are blunt instruments that tend to do different things over time, so they're not tools. They're ways of thinking, they're ways of knowing, they show at times interesting correlates, they can provoke conversation, you can use them as text, you can use them for reflection, but they're not a tool, at least not in the way a tool would be traditionally defined.

I think this description offers an interesting way to consider the Progress Reports and Quality Reviews as they were used by networks to develop shared plans. They were used by some networks as "text" that provoked a conversation and it was up to the people to negotiate meaning to develop their shared plans of support.

Influences on the use of Progress Reports and Quality Reviews

With my final research question, I sought to understand what influenced the ways in which networks' used the Progress Reports and Quality Reviews in their efforts to support school improvement. This question was presumptive given that past research and theory suggest that it was improbable that networks would enter into productive relationships with schools and use accountability tools as a means for improvement. While I did find evidence of principals buffering against network intrusion, there was also evidence, as described above, that networks were able to engage schools in active, relevant use of the tools in order to develop capabilities and support structures for improvement. From interviews and observations, I found that there were three central factors that influenced the quality of partnerships and use of the tools in the networks' partnerships including interpersonal relationships, talent, and role conflict. Below, I offer a discussion on each of these factors and how they influenced networks' use of the tools.

Interpersonal relationships

Past relationships mattered a great deal to the depth of engagement and health of network's partnerships with schools. This finding really builds on other research that found that relational trust was a positive resource in school improvement efforts. Bryk et al (2010) wrote, "Relationships are the lifeblood of activity in a school community. The patterns of exchanges established here and the meanings that individuals draw from these interactions can have profound consequences on the operations of schools, especially in times that call for change" (pg. 137). These authors suggest that trust is grounded in respect. When networks worked with principals whom they did not respect, the use of the tools was not particularly helpful in generating supports, and vice versa.

I looked closely at two schools that were being supported by CFN 706. The principals at both schools were in their first year as principal. The schools were identified as persistently low performing, which meant that they had been flagged by central and required to create a Targeted Action Plan with their networks. In 2010, the DOE started to require schools that had been identified with quantitative and qualitative data to be the 15% lowest-performing schools to develop a Targeted Action Plan (TAP) with their networks. TAP plans required the school and network to identify formally the

key areas in which the school needs to improve and articulate how the network would support these areas. Central office monitored TAP plans – requiring networks to submit status reports at three points in the school year. According to Parthenon (2013), "three-quarters of the schools with a TAP that received a D or F on the progress report in 2011-12 saw an increase of at least one letter grade in 2012-13; 38 percent increased their results by two or more letter grades" (pg. 10). TAP plans required the networks to work closely with schools to collaborate, but this top-down requirement was not as powerful as the social dynamics between networks and schools as evidenced by the stark difference in the relationship CFN 706 had with these two schools.

Over the past three years, P.S. 373 had seen a steady decline in its Progress Report grades – going from an A in 2010-2011 to D the next two years. The school got a Developing on its most recent Quality Review (second from the bottom). In 2013, the school got a new principal, Mrs. B. Mrs. B. had a good relationship with CFN 706 because she had worked with its coaches and network leader at her previous school, where she worked as an AP. When I met with her in the spring of 2014, she raved about her relationship with the network and credited the coaches with helping her with the teacher teams and budget issues at P.S. 373. She described relying on the network to help her "move the teachers and classroom practice." I observed several teacher team meetings that were facilitated by the network coach; the principal joined several of the sessions. The principal discussed her relationship with the network with me in front of the coach saying, "There's nothing I am going to say you can't or shouldn't hear." She spoke very positively of her experience with the network's help with the challenges she faces in her building. At the end of the meeting, she invited the network's ELA coach to the school's summer retreat.

The story at P.S. 1672 was drastically different. P.S. 1672 in the Bronx earned a C for three consecutive years and had been targeted for additional interventions. The principal, Mr. Angel joined midway through the prior year. Like Mrs. B., he inherited the network, CFN 706, but unlike Mrs. B., he no prior experience with the team and was not familiar with them. When I met with him on the last day of the school year in June, he was quick to criticize the network's efforts to support P.S. 1672. He described his school as a bleeding patient that the network abandoned; he saw himself as the

emergency room triage nurse – "trying to stop the bleeding before the doctor arrives," he explained. At several points, he insinuated that the network's negligence was "criminal." He believed that the network failed to step in when there were clearly "red flags" including the lack of a PTA and SLT at the school (also "criminal" in his description as these are required by law). He felt the network not only failed him and the school, but he held the opinion that they intentionally tried to block his efforts to make change at his school – citing a budget request that had been denied as evidence. Mr. Angel elected to leave CFN 706 and chose to affiliate with a different network. Mr. Angel had relationships with several principals in the new network, and had worked with several of the new team's coaches in his former role as an AP before taking over as principal at P.S. 1672. Mr. Angel and Mrs. B. were not exceptions; rather, it was really common to hear references to past relationships and former affiliations (and allegiances) as critical influences on how productive a network could be with a school.

Talent

More than eight hundred people worked on network teams across the city. Given the size of this pool of workers, there were bound to be variations in the level of knowledge, experience, and talent within and across teams. This variation influenced the kinds of support and the quality of support a network team provided to its schools. As one Cluster leader shared, "not every network has five strong coaches. Some of them are watered down. Not all have five instructional folks. Some are heavier on the ops (operations) side. It's difficult." He further explained, "Talent is not always there. So, you try to make matches to what the need is there. Not all of our people need the same stuff. So, it's about trying to get people - not just smart people, but smarter about the people we put in places. Sometimes, we make matches because it's the smartest person in the room, but smart doesn't always transfer."   The kind, amount, and depth of talent on a team influenced how networks provided supports – including how specialized a team could be in helping schools meet technical or adaptive challenges.

The network leaders had to serve as managers of their teams, and "pick their battles," as one network leader explained. He elaborated, "I can't speak for all, but, at least in my case, it created a challenge around how does a network support a network of

principals when we only have one kind of achievement coach, or three kinds of achievement coaches? We don't have enough people to support all of the needs of all of the schools. So you, pick your battles."

Picking battles was what some network team members needed to be able to do – especially in the instances when network members needed to be able to have 'hard conversations' and convince principals to do work that they may not want to be able to do. For example, one network leader explained,

> You have to be able to have a tough conversation. If you get into a school, and the principal is having you sit in his office and spending three hours talking to you and griping, and not giving your team the opportunity to get into classrooms, to work with assistant principals, to work with coaches in the school, and to really find ways to make sure the work is landing and that you're building capacity, you have to be able to say to a principal, 'look, I can sit here on a different day, but this is what I came here to do today and this what really have to get done. And that's the problem, we have a lot of, I'm not saying this for the cluster, but city-wide, we have a lot of young, inexperienced coaches that really aren't comfortable having those difficult conversations with principals so they end up sitting in the principal's office doing their CEP (comprehensive education plan) for them when they really need to working in the classrooms with the teacher.

Several times I heard, "not all coaches are created equal." One Cluster team member commented on this point when she explained, "And then there's quality. I mean you can anticipate those when you're looking at the Progress Report that not everybody speaks data, and not every achievement coach speaks data, which is unfortunate, but it's true. So, how do you teach achievement coaches to speak data when they don't? I once watched an achievement coach walk into a school and completely misread the progress report data that the principal and I had to go back and like, how do you say to a colleague, you got it wrong? And so, if I observed it once, how many times did it happen? I don't know. Maybe that was the one and only unique situation, but I find that hard to believe." The uneven talent in the system constrained some teams and presented additional challenges to providing flexible, adaptive support to schools. This speaks to other research that has highlighted the limitations of formal roles and structures, or

188

artifacts at changing practice. All of these things depend upon the depth of knowledge and know-how that people possess as they enact practice or live their performances, and as such, people need opportunities to learn how to use new artifacts or operate within new structures in order for those things to make a difference in practice (Coburn & Stein, 2008; Pentland and Feldman, 2008). As Sykes, Ford and O'Day (2009) highlight, "if as a nation we are to significantly improve instruction and student outcomes, then adults in the system must not only be motivated to change what they do (e.g., through policy pressure), but they must also learn to do their work in different and more productive ways" (pg. 772).

Role conflict

A general theme that emerged from interviews and observations was the conflict that the networks faced as they tried to balance demands from central office to move schools on the Progress Reports and Quality Reviews with demands from principals; I referred to this as role conflict. Role conflict did not always exist for all network leaders; rather, it depended on the specific dynamics in the relationships between principals and networks. Depending on the principal and the school, the set of demands on the network team could be congruent, in the case that a principal's goals aligned with central's goals, or conflicting, if a principal wanted to buffer his school from pressure from central office. In the instances when a principal rejected the principles, or legitimacy, of the Progress Reports and Quality Reviews, the networks were in the difficult position of having to serve as a partner to the principal and persuade him that these things mattered – as it would matter for the network's rating. What complicated matters further was that networks had no authority to tell principals what to do; they had to find ways to manage productively the tensions that arose when a principal's goals were not congruent with the DOE's strategy for school improvement.

This point was made earlier when I presented some of the struggles that networks faced when they tried to provide targeted, site-based supports. I turn again to the network leader who explained that "…the principal has to be on-board and they have to ensure that they're putting the structures in place for this to happen." The network leader shared that she could be frustrated by the fact that "…as a network leader, you are not a

supervisor. You cannot tell a principal what to do or how to do it. You have to really try your best to develop a good strong relationship with the principal and really try to support their work, but there have been occasions where I have suggested certain curriculum that they should use or programs or scheduling or so on and when I've gone back to see that those things were implemented, they weren't… and that's when I become frustrated sometimes, as a network leader."

The networks were held accountable by the DOE and subjected to evaluation under the Network Performance Management System; NPM was used to rate the networks according to how well their schools performed on the Progress Reports and Quality Reviews. Getting a school to move, could be uncomfortable work for the schools, which, at times conflicted with a principal's desire to feel safe, comfortable, and protected by the network. Networks had to find ways to navigate potential tensions. For example, one network leader described a time when her team was helping a school prepare for a Developing Quality Review and this tension arose. She shared,

> You go to the one-day review. It's a little challenging, I'll say, because these are people that are paying you, my team for support, and many times, they sit there, they rate themselves…they go and sit in their room, we sit in a room, we score, we come back and say, 'okay, 1.1, what do you think you are?' And they'll say, 'well-developed.' And then the network will say, 'developing.' And there you go: conflict. And at one school, I'm sure that's why she left. I am almost positive because we went back four times because they were insisting they were 'well-developed.' The second time, the place was filled with stuff - folders, documents, ridiculous. I had to tell them, 'stop. Don't bring me anymore in here because these documents are digging you worse into the hole because it's a show, and sorry, it's not a good artifact. So, it can be tension - especially, if people aren't honest and reflective about the work that is in front of them.

As this quote suggests, whether a tension could be productive or not, depended on the school's view of the accountability tools. If the school treated them as compliance tasks, it made it difficult for networks to treat them as learning tools and posed a challenge to the network as it sought to help the school develop organizational learning capabilities. Networks were not supplanting superintendents, who in the past, were looking for

compliance. Rather, the networks were meant to be support partners – helping schools develop sustainable continuous improvement practices, but this became complicated, if schools continued to frame accountability, and the networks, as compliance-based interactions. The degree to which the networks and Progress Reports and Quality Reviews could be influential was completely dependent on a school leaders' attitude towards the tools and the sense they made of them. As brokers, the networks could develop resources and try to create sensemaking opportunities to jointly interpret the Progress Reports and Quality Reviews, but they were confined by how open a school was to engage in such opportunities.

One of the original architects of the accountability system likened the idea for the networks to agricultural extension service from the early 20th century, which "turned out to be an amazingly important to the 20th century power of the United States as an agricultural power because they facilitated getting through the Dust Bowl and dealing with all sorts of issues that farmers face, but they didn't do it cause they had no authority over farmers, all they could do was help them," as he explained. The extension service was created in 1914 through the Smith-Lever Act; the mission of the extension service was to bring evidence-based research and ideas to farmers and rural communities to promote positive changes through collaboration (USDA). The source of authority for the extension service was knowledge and expertise and an ability to collaborate with farmers in the service of improving farmers' situations. However, the analogy fails to account for the motivation of the farmers to engage with and seek the help of the extension service in the first place. The farmers, unlike the principals in New York City, operated in a market where better crops would lead to more profit. While Mayor Bloomberg and Chancellor Klein were trying to expand the number of quality options available for parents to choose from, there were limited market pressures on many schools, and principals' motives may not have been the same as Dust Bowl farmers.[6]

Despite the limitations of the analogy, according to the architect, "this was the model" they were going for with the designs for how the networks were meant to work. Although, looking back on the reforms and how they were designed, he expressed, "the biggest thing that we missed was the networks were wrong and the networks were largely

---

[6] I am grateful for David Cohen's help in seeing this point.

about service to the schools, the networks were supposed to be both about service to the schools and about providing a facilitative process to help their schools improve for which they felt accountable." He continued to suggest that they got it wrong because, "the way we set it up and the schools choose them and sort of some other things, it became, 'how well do you serve your schools?' And so the schools wanted you to get stuff out of their hair – do special ed assignments and admissions and food and transportation – all that junk, get as much of that off the school as possible, but the schools also didn't want anybody looking over their should about how well their kids were doing." He concluded by saying, "we missed a beat there." The beat they missed was figuring out the right resources that were needed to make principals think like managers and begin to see their role as managing for improvement. The networks were supposed to be instrumental in the management of improvement – not just getting things out of principals' 'hair' or 'off their plate'; the point was to trigger authentic commitments to continuous improvement practices.

Across interviews, the distinction between "service" and improvement-focused work was characterized as "taking the work seriously." To take the work seriously meant to buy into the principles of the Quality Review and the Progress Reports and see these tools as providing a principal with valuable feedback that could be used in the service of school improvement. The thinking was that the network teams, because they were not supervisors, would be positioned to support and facilitate improvement through helping the schools see the evidence and learn from it versus past models where accountability was associated with compliance-based tasks mandated by central office. However, helping principals make this mental shift – to see the tools as learning tools, not compliance based required that the networks figure out how to differentiate collaborative instruction and facilitation from mandates from some external agent (central office). But as one senior official explained, "this is tricky business."

The senior official also offered that the success of whether or not the Quality Review became an "improvement tool" could be determined, "If the network team takes that day review very seriously, and has, and this is tricky business, where you have standing with the principal and the school community, where you can go in for a day and

give some hard feedback – not just glowing feedback and manage to build on that, then you have a real tool that's an improvement tool."

Yet, he was cautious. He said what others alluded to: "There's so many ways for us to fall down along the way." He provided an example saying, "They [networks] can say, I am going to take this seriously, but I am afraid of this principal. If I give the principal bad feedback, then they don't let me back in the building, you know, to do the work that I have already been doing." The networks had to manage their relationships with principals and push the work without pushing too far such that a principal would shut her door.

Balancing principals' demands with the DOE's was really difficult for some networks that wanted to please principals. Strong principals, who often led really good schools, knew what they wanted and would demand it of the networks. Networks would like to comply, but they did not have unlimited resources and had to make decisions about allocating coaches and determine where they could get the most leverage to advance the network and their schools, but this was not easy for the networks to manage on their own. For example, one cluster leader explained, "We too often see the strongest coach that the data shows was able to move schools - was able to move the needle on student achievement side and then all of a sudden, they are with the three strongest schools in the network…the principals know that they are the strong and the principals know what they need and so they say, 'no, I want this one.'" In this cluster, the Cluster leader made it a habit to "get involved" in these situations to "see the matches and we lean on them to make changes." However, not all networks had this kind of Cluster leader who knew his coaches and schools so well. In many ways, the networks had to find ways to make matches that fit with the principals' wants and balance that with the needs across the network given the team's constraints and capacities.

What allowed some networks to successfully navigate the tensions – without Cluster interventions – was what one deputy chancellor described as "earned authority." He explained that, "…the people that I view as the most successful within this have also the personality and/or the skill-set to gain what I call earned authority in terms of, our strongest network leaders are able to go and tell a principal, 'you should do this, you should do that, you should do this.'" According to him, the network leaders who possess

this "skillset" acquire it through two things, "(a) it's coming from a place of the best interest of the kids in the school – not because I just think it's the right thing to do, and (b) the principals receive it knowing that the network leader is someone they respected tremendously and that they're looking out for the best interest – they're accountable for the outcomes of their schools as, that isn't just a self-interest or personal interest or power trip, but it's about the best interest of the school." It mattered whether principals recognized that networks, too, were accountable for outcomes, and if the shared accountability was recognized.

The networks were able to gain "earned authority" by using the Progress Reports and Quality Reviews as evidence to explain to a school, "I think you should do this, and here's why; you don't have to do it, but this is why I think you should," as explained by one deputy chancellor. He offered, "And so, I think the ability to go and have a leadership style that is not hierarchical or bureaucratic or top down, but being one of collaboration and like being deeply invested and feeling deeply accountable for the product and the outcomes enables that conversation."

However, collaboration was not always achieved in the service of school improvement, but were necessary to maintain a service relationship with the school. For example, one network leader explained, "

> …there's planned things, so those planned things and you knew going in what you were doing and they were scheduled and there were the everyday things that all of a sudden there was a problem, a 311 call with a parent at a school and I had to go rushing, so a lot of those kinds of things as a network leader too, a lot of putting out fires, which takes you away from some of the instructional initiatives that you're working on. That's why I did have a deputy in charge of instruction who could make sure, I mean I was always part of it, always sat on the planning meetings cause that is my love and I think my expertise, but there were many times like I couldn't because I had to run to put out fires.

In these kinds of instances, the networks had to appear like they were 'helping,' but the version of help was not necessarily focused on moving the school towards continuous improvement practices.

The networks were intended to enable school improvement through their facilitation of the use of the accountability tools with their schools; they were support to balance accountability. One senior executive explained, "The networks have been around as long as the Quality Review has. And basically, they're two, maybe, complementary roles, accountability on one hand and support on the other hand. So they actually don't have to be confused with each other, hopefully, they complement each other, the QR and the accountability measure and then the network person should be sitting on that final feedback or final session so that they can listen to the outcome along with the school leaders and then go from there to help school leaders think about what the next steps are based on the feedback they just heard."

Because the networks, as one architect explained, "felt the same responsibility that you were describing for the schools on themselves, but felt that they had to kind of earn their authority by helping, by proving that they could help the schools get where they needed to go rather than having been given the authority from the Central bureaucracy to assure, order, or demand compliance with whatever it was." The networks needed to prove that they could help schools get where they needed to go, but the problem was that not all schools agreed to where they needed to go, and that was going to take, as one deputy chancellor expressed, "culture change and evolution" to see accountability as a learning opportunity.

Summary and Conclusions

The purpose of this article was to present how the networks in New York City used the Progress Reports and Quality Reviews in the service of school improvement. I began by explaining that the conditions in which these networks were designed, and the context in which they operated. The networks' origins and the environment where they worked were fairly unconventional when compared to other district level reforms and foreign to those who had long careers in the NYC system. Mayor Bloomberg and Chancellor Klein, through the Children First Reforms, completely changed school governance, school accountability, and supports for schools in NYC.

These reforms changed the relationships between central office and schools – by shifting responsibilities, central office was meant to become a manager of a portfolio of schools, and principals were supposed to became CEO's of their schools responsible for

managing their school's improvement efforts (Wohlstetter, Smith and Gallagher, 2013). NYC's Department of Education (DOE), created network teams to help principals adjust to autonomy and adapt to new expectations for performance under the portfolio management structure, where they had to prove that they could hack it, else they would be replaced by a school that could.

The DOE's demands required schools to develop capabilities to function as autonomous schools equipped to coordinate organizational conditions to improve instruction and ensure more students made progress regardless of past performances or neighborhood conditions.   Schools needed to be able to understand and respond to the Progress Reports and Quality Reviews; networks were provided to help schools develop the necessary capabilities to function autonomously and respond to the new kind of accountability tools. These new tools, the Progress Reports and Quality Reviews, were used to hold schools accountable; there were swift and steep consequences for schools, if their grades on the Progress Reports and ratings on the Quality Review remained low.

The networks, too, were subjected to strict accountability for performance under a public-sector market arrangement. The public-sector market involved two forms of accountability for networks. The first was a bureaucratic: networks were rated by central office using a Network Performance Management system that ranked networks based on how well their schools performed on the Progress Reports and Quality Reviews. Principals could choose which network they wanted to affiliate; each spring, they were given an opportunity to re-affiliate, if they wanted to leave a network and join another one. This created a market-place in which networks competed to serve schools and satisfy principals.

In this study, I found that when there was congruency between principals' and central's goals, networks were in a much better position to provide productive support aligned with the Progress Reports and Quality Reviews; it was easier to broker commitments and facilitate strong participation in partnerships. Yet, the times when a principal's goals were not congruent with central office's goals, the network had to manage this tension. In some instances, the Cluster could serve as a mediating agent, but other times, the network risked losing a principal to the marketplace. To address the tension, the networks would either provide alternative services that were not necessarily

supports for school improvement, or they would find ways to have hard conversations with principals that were meant to convince them of the merits of focusing on the accountability tools as methods for improving the school's ratings. Whether these efforts touched instruction depended on the network – its legitimacy, relevance, helpfulness, and friendliness in the eyes of the principals.

This analysis has several implications for the practice and research on the use of external support for school improvement in portfolio managed districts. First, this research suggests that balancing support and strict accountability encouraged people to take the work of responding to accountability "seriously." Many of the networks took the work seriously because there was shared accountability between the networks and schools, and networks became formally responsible for some parts of the Quality Review encouraging closer ties to the work. For example, one cluster team member explained,

> because it's (Quality Review and Progress Reports) part of the network accountability, they pay closer attention and I do think that they organize their coaching support to make sure that systems and structures in schools are strengthened. Yeah, we now, almost every network now does mock quality reviews to varying degrees and to varying organizational success, but they all pay attention to that, they all support their principals in preparing for the quality review. Now that they're all doing like the DQRs and the PFSQRs, whatever they are, they're even more invested in getting the process right, getting the feedback right, making sure that it's a meaningful investment of time for both them and the schools.

The bi-level accountability scheme encouraged greater emphasis among networks on getting it right with schools. This was not a voluntary learning tool; rather, it was something the networks and schools learned together to use in ways that strengthened the "systems and structures in schools," according to the team member.

Second, this study extends findings on the role of authority and status in external support partnerships. In examining partnerships between a university-based research center and schools, Coburn, Bae and Turner (2008) found that status issues that privilege academic knowledge posed challenges to partnerships between schools and the center; it was difficult for these partners to establish "normative agreement about appropriate

authority relations" (pg. 40). Authority matters in these partnerships, as these authors suggest, because, "the development of clear authority relations actually enables productive working relationships. Shared understanding of appropriate roles and relationships provides guidance for interaction and decision making, and for mitigating against breaches, power struggles, and misunderstanding" (Coburn, Bae, Turner, 2008, pg. 41). The networks in NYC had neither status, nor formal authority, and had to earn authority and prove to schools and central office that they could offer value to schools. In many cases, but not all, their use of the accountability tools to target and develop worthwhile supports allowed them to earn the necessary authority to move schools in particular directions.

Third, this research highlights the role of accountability tools as levers for support. Most networks, not all, used the combination of the Progress Reports and Quality Reviews to prioritize their work and manage the uncertainty of supporting diverse schools. The Progress Report was less valuable to networks as they designed supports because as one network leader explained, "everyone knows its not a good measure because you can walk into an A school, but if the instruction is not good, then it's not a good school. Similarly, you can walk into a C or D school and the principal is doing all the right things…teacher teams, curriculum, and other innovations – we take the grade with a grain of salt." By contrast, the Quality Review rubric became a valuable guide for developing network-wide professional development, as well as, targeted and specific site-based supports for developing school-wide capabilities for improvement.

Lastly, this research suggests that using networks and accountability tools as levers for school improvement takes time to build the capabilities to use them well. While there were instances where networks and schools would try to find quick fixes like taking kids to summer school to improve scores on the Progress Reports, the harder, more adaptive challenges of using the Quality Review rubric to change instructional coherence and systems for improving instruction took time. I heard reports that it made a significance difference when people took the long-view and used the Quality Review to develop continuous improvement practices. Although, it took time for this to develop. For example, one cluster team member shared,

So, the one thing I think changed significantly is the development of teacher teams and really understanding that teacher teams are a major lever for moving the work forward. Like, I can confidently say, six years ago, not every school had teacher teams. There were some schools that had them – some that had other than department teams, like no teams. So now I feel like all schools have teacher teams and I think that the work that teacher teams are focusing on is much more meaningful than it was really prior to the quality review but has strengthened as a result of the quality review… That didn't always exist in our schools and I really think the quality review has really pushed that forward.

Research supported to the point that taking the long view when designing development opportunities and supports paid off over time. For example, one study revealed: "Better outcomes on the Quality Review are associated with better outcomes on subsequent Progress Reports, as well as, with better postsecondary outcomes for students" (NYC DOE, 2013 What's next). Yet, as a cluster team member expressed, "So I think it's both qualitative and quantitative, the reasons why schools are paying more attention to it…but, I think that people are – it took a while to see the value of this."

For researchers, this essay suggests the importance of extending descriptions of the effects of combining management principles like continuous improvement and instructional improvements. Such research should consider whether schools that lack access to an instructional system are able to implement and sustain continuous improvement practices independent of help from an external support provider, or should we expect that schools will always need external help? If help is necessary for sustaining commitments, what are the implications for accountability policies that aim for improvement even across schools?

This research focused on the networks without really understanding the effects support had inside schools. Future studies could consider the effects of bi-level accountability schemes on the day-to-day practices inside schools. Do these systems lead to better supports for school improvement? If so, under what conditions and for whom are the supports better? What are the results?

Finally, accountability uses student achievement to gauge whether accountability worked as intended, but do students have better educational experiences inside schools

that embraced the Quality Review and Progress Reports as means for improvement? Did one tool have more influence? If so, for whom and under what conditions? What were the results? The answers to these questions could lead to broader encouraging use of Quality Reviews as evaluations attached to stakes, or they may suggest that they are not worth the time or money or significant effort it takes to implement something like this as an accountability tool. In New York City, the Quality Reviews appeared to be worth it. One of the cluster team members shared, "a couple years back, when Quality Reviews were on the table – about four or five years ago, cause we were going through a massive budget cut, Joel Klein came and said, I don't know. It costs something like $21 million to run that office, and he's like, 'it's the best $21 million we have spent.'" More research is needed to know its full value.

References

Booher-Jennings, J. (2005). Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal.* Vol. 42, no. 2, pg. 231-268.

Bryk, A.S., Sebring, P.B., Allensworth, E., Luppescu, S., Easton, J.Q. (2010). Organizing schools for improvement: Lessons from Chicago. The University of Chicago Press: Chicago, IL.

Childress, S. Higgins, M., Isimaru, A., Takahaski, S. (2010) Managing for results at the New York City Department of Education. In Education Reform in New York City Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Harvard Education Press. Cambridge, MA.

Coburn, C. E., Bae, S., Turner, E.O. (2008). Authority, status, and the dynamics of insider-outsider partnerships at the district level. *Peabody Journal of Education*, vol. 83, no. 3. Pg. 364-399.

Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality?. The Teachers College Record, 106(6), 1145-1176.

Elmore, R. F. (2004a). School Reform from the Inside Out: Policy, practice, and performance. Harvard Education Press. Cambridge, MA.

Elmore, R.F. (2004b). Conclusion: The problem of stakes performance-based accountability systems. In Redesigning Accountability Systems for Education. Susan H. Fuhrman and Richard F. Elmore (eds). Teachers College Press. New York, NY.

Elmore, R.F., Abelmann, C.H., Fuhrman, S.H. (1996). The New Accountability in State Education Reform: From Process to Performance. In Holding Schools Accountable: Performance-Based Reform in Education Helen Ladd (ed). The Brookings Institution: Washington, D.C.

Finnigan, K.S., Bitter, C., O'Day, J. (2009). Improving low-performing schools through external assistance: Lessons from Chicago and California. *Education Policy Analysis Archives*. 17(7).

Firestone, W.A., Gonzalez, R.A. (2013). Educational tug-of-war: internal and external accountability of principals in varied contexts. *Journal of Educational Administration*. Vol. 51, no. 3, pg. 383-406.

Heifetz, R.A., Linsky, M., Grashow, A. (2009). The practice of adaptive leadership: Tools and tactics for changing your organization and the world. Harvard Business Press: Cambridge, MA.

Hemphill, C., Nauer, K. (2010). Managing by the numbers: empowerment and accountability in New York City. Report from The Center for New York City Affairs, June 2010.

Hill, P. T. (2010). Leadership and Governance in New York City School Reform. In Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Education Reform in New York City. Harvard Education Press. Cambridge, MA, pg. 17-32.

Honig, M.I. (2006). Building Policy from Practice: Implementation as Organizational Learning. In Meredith I. Honig (ed). New Directions in Education Policy

Implementation: Confronting Complexity. State University of New York Press: Albany, NY. Pg. 125-148.

Honig, M.I., Copland, M.A., Rainey, L., Lorton, J.A., Newton, M. (2010). Central office transformation for district-wide teaching and learning improvement. Center for the study of teaching and policy at the University of Washington. Report for The Wallace Foundation, April 2010.

Honig, M. I., DeArmond, M. (2010). Where's the Management in Portfolio Management? In Katrina Buckley, Jeffrey Henig, Henry Levin (eds). *Between Public and Private: Politics, Governance, and the New Portfolio Models for Urban School Reform*. Harvard Education Press: Cambridge, MA. Pg. 195-216.

Ikemoto, G.S., Marsh, J.A. (2007). Cutting through the "Data-Driven" Mantra: Different Conceptions of Data-Driven Decision Making. In Pamela A. Moss (ed). *Evidence and Decision Making*. The Yearbook for the National Society for the Study of Education, Part 1. Blackwell Publishing: Malden, MA. Pg. 105-131.

Kane, T. J., Staiger, D.O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*. Vol. 16, no. 4. Pg. 91-114.

Klein, J. (2014). Lessons of Hope: How to fix our schools. Harper Collins: NY, NY.

Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*. Vol. 29, no. 2., pg. 4-16.

Marsh, J. A., Strunk, K.O., Bush, S. (2013). Portfolio district reform meets schools turnaround. *Journal of Education Administration*. Vol., 51, issue 4, pg. 498-527.

Maxwell, J.A. (2005). *Qualitative Research Design: An Interactive Approach*. Sage Publications (2nd ed.). Thousand Oaks, CA.

Miles, M., & Huberman, M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher, 13*, 20–30.

Mintrop, H., Trujillo, T. (2007). Practical Relevance of Accountability Systems for School Descriptive Analysis of California Schools. *Educational Evaluation and Policy Analysis.* Vol. 29, no. 4, pg. 319-352.

Murnane, R.J. (1987). Improving education indicators and economic indicators: the same problems? *Educational Evaluation and Policy Analysis*, 9, 101-116.

Nadelstern, E. (2013). 10 Lessons from New York City Schools: What really works to improve education. Teachers College Press: New York, NY.

Nadelstern, E. (2012). The evolution of school support networks in New York City. Center on Reinventing Public Education (CRPE) Working Papers 2012 #2. Retrieved: www.crpe.org.

Newmann, F.M., King, M.B., Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. Harvard Educational Review. 67, 1; pg. 41-65.

New York City Department of Education (2013a) What's Next for School Accountability in New York City? Released November 2013.

O'Day, J. A. (2002). Complexity, accountability, and school improvement. In Redesigning Accountability Systems for Education. Susan H. Fuhrman and Richard F. Elmore (eds). Teachers College Press. New York, NY.

O'Day, J., Bitter, C. (2010). Improving Instruction in New York City Schools: An

Evolving Strategy. In Jennifer O'Day, Catherine Bitter, Louis Gomez (eds). Education Reform in New York City. Harvard Education Press. Cambridge, MA.

Parthenon (2013). An Assessment of the New York City Department of Education School Support Structure. Report November 2013.

Pentland, B.T., Feldman, M.S. (2008). Designing routines: On the folly of designing artifacts, while hoping for patterns of action. *Information and Organization*. 18: 235-250.

Sabel, C. (1993). Learning by monitoring: The Institutions of economic development. Working Paper #102 August 24, 1993.
Retrieved:
http://www3.law.columbia.edu/sabel/papers/Learning%20by%20Monitoring.pdf.

Shipps, D. (2011). Empowered or beleaguered? Principals' accountability under New York City's Diverse Provider Regime. *Education Policy Analysis Archives.* Vol. 20 (1).

Smylie, M., Corcoran, T. (2006). Nonprofit organizations and the promotion of evidence based practice. Paper presented at the annual meeting of the American Educational Research Association.

Spillane, J.P., Diamond, J.B., Burch, P., Hallet, T., Jita, L., and Zoltners, J. (2002). Managing in the Middle: School leaders and the enactment of accountability policy. *Educational Policy*. Vol. 16, no. 5. Pg. 731-762.

Supovitz, J. A. (2006). The case for district-based reform: Leading, building, and sustaining school improvement. Cambridge, MA: Harvard Education Press.

Stein, M.K., Coburn, C.E. (2008). Architectures for Learning: A comparative analysis of two urban school districts. *American Journal of Education.* Vol. 114, no. 4, pg. 583-626.

Sykes, G., O'Day, J., Ford, T.G. (2009). The District Role in Instructional Improvement. In Gary Sykes, Barbara Schneider, David Plank (eds). *Handbook of Education Policy Research*. Routledge: New York, NY. Pg. 767-787.

United States Department of Agriculture, National Institute of Food and Agriculture. Topic: extension. Retrieved: http://nifa.usda.gov/extension. May 21, 2015.

Wohlstetter, P., Smith, J., Gallagher, A. (2013). New York City's Children First Networks: Turning accountability on its head. *Journal of Educational Administration.* Vol. 51, no. 4, pg. 528-549.

APPENDIX E: NYC School Support Organizational Chart
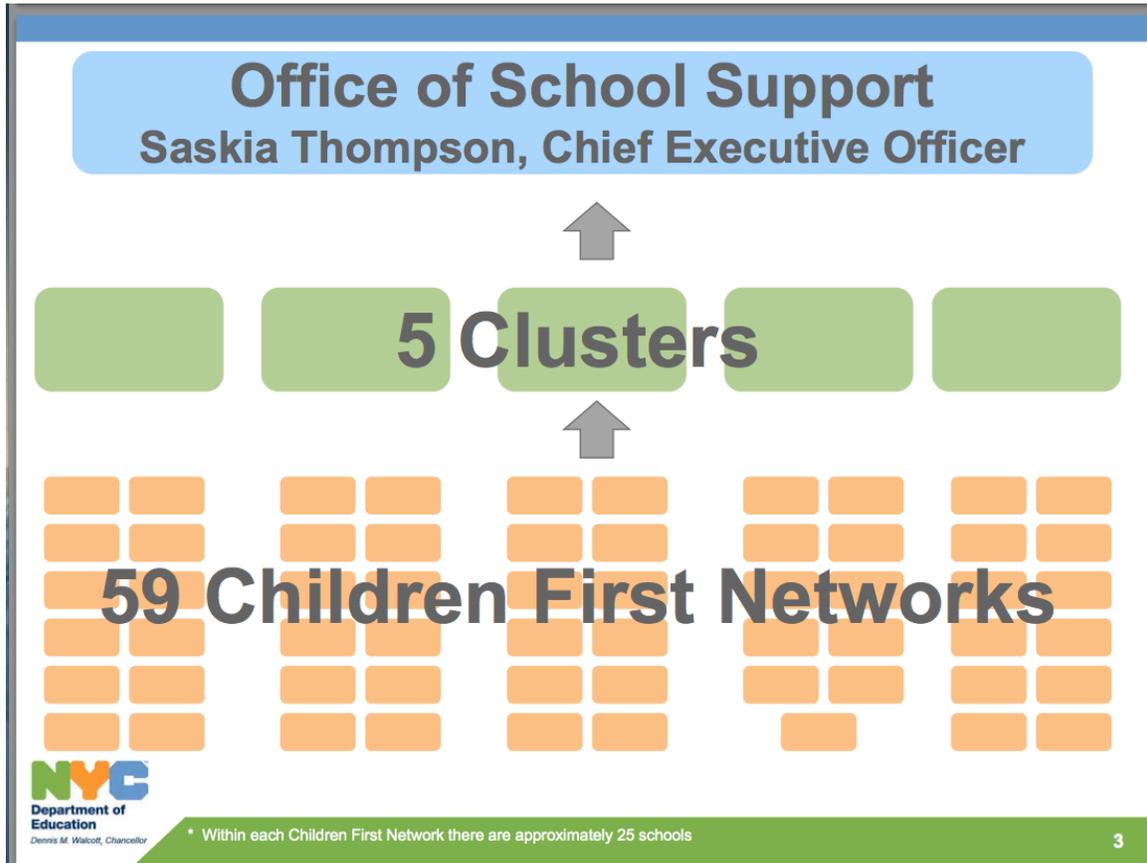
Source: NYC DOE, NYC DOE Structure for Supporting Schools, December 2011, pg. 3.

CHAPTER V

CONCLUSION

"Accountability is essential. It tells people where they stand and, if implemented effectively, shows them how to improve. When necessary, accountability enables them to be removed if they don't meet standards. No one likes to be monitored, evaluated, or corrected. Accountability can have painful dimensions, no matter how it's done, a lesson I personally learned whenever the mayor called me on the carpet for, in his words, 'screwing things up.' But we couldn't improve academic outcomes, if we didn't improve teaching, and this would require evaluating teachers, enhancing their strengths, addressing their weaknesses, and removing those who weren't qualified."

- Joel Klein, 2014

The three essays in this dissertation were intended to contribute to discussions about school accountability as a way to improve schools. Conventional thinking about accountability is that it tells us about our schools and helps ensure that public schools are living up to our aspirations and if a school is not, then accountability shows them how to improve. This logic suggests that because all schools are held accountable, then accountability *should* help all schools improve. While some schools have been able to respond to monitoring and evaluations by correcting mistakes when necessary, many, many schools have not. As a system-level reform, the results from instituting strict accountability standards and assessments have been mixed and uneven. I think we could do better.

     With this dissertation, I called for a reconsideration of the taken-for-granted view that if we just hold schools accountable, they will improve. Based on my experiences as a teacher in New York City during Mayor Bloomberg's takeover of the city's schools, I sought to understand better the novel accountability system that was implemented during the Mayor's overhaul of the NYC's school system. What appeared to some to be another case of educational upheaval appeared to me to be an unprecedented attempt by a district to balance pressure and support for improvements in teaching and learning.

With their Children First Reforms, Mayor Bloomberg and Chancellor Klein sought to transform an entrenched and slow-to-change system into a dynamic, diverse and innovative "system of great schools" (Klein, 2014) in which continuous improvement and Total Quality Management became regular practice. As one of the former executives expressed, calling the DOE a "dynamic environment" is "a strange thing to say about the largest bureaucracy in New York City," but that was what they were going for. They intentionally tried to avoid conventional educational reform tactics.

School accountability was a central pillar of Children First, but it was one reform of many – and as such, telling the story of accountability reform that occurred under Mayor Bloomberg is inherently entwined with several other system-level initiatives including curriculum reforms, the development of talent pipelines, redistribution of school funding, changes in governance, and creation of networked school support organizations. Accountability was a superstructure for the system Bloomberg and Klein were hoping to construct. While I tried to attend to the composite of changes that took place in NYC, as one researcher, I was fairly limited in my reach. As a result, this dissertation begins conversations and shows possibilities. Those seeking exact instructions for replicating New York City's approach may be disappointed.

In Chapters Three and Four of this dissertation, I focused specifically on the design and use of two accountability tools in New York City: the Progress Reports and Quality Reviews. I explored these tools for three reasons. The first was because when they were introduced, the Progress Reports and Quality Reviews were unlike any other form of school accountability in the U.S. Second, these were part of an unconventional set of system-level reforms that included the creation of novel support organizations intended to bolster learning and development among those in schools to be able to respond to the novel accountability that included them being held accountable for results, as well. Traditionally, it is rare within district reform that support providers would be held directly accountable for school outcomes. Yet, NYC created a system, the Network Performance Management System, to do just that.

Lastly, NYC's experiments were conducted at a scale and pace rarely seen in education reform. The New York City Department of Education manages more schools than twenty-eight states in the U.S. (Parthenon, 2013). While, the findings from these

studies are not necessarily generalizable, they certainly offer some useful lessons for the use of accountability as a system-wide intervention.   There are two broad themes that emerged from my exploration of New York City's accountability system that could prove useful to scholars, policy makers, or educators. These themes reflect tensions that were managed through the design and use of the Progress Reports and Quality Reviews including a tension that between focusing on internal versus external environments and, bold versus incremental change. Below, I highlight the themes and close with a summary.

External and internal environments

Accountability policies involve a tension for schools that can seem, at times, at odds. Their external environments, issue the rules and instruments that evaluate schools and judge them as to whether or not they are adequate. The external environment places demands and applies pressure on schools for improved outcomes. Schools internal environments, those in which teachers access and create norms, routines and administrators enforce standard operating procedures influence the day-to-day work, as well as, provide whatever resources a school may possess to respond to external pressures for improved performance. Yet, depending on the aims and designs created in the external environment, a gap may be created between what any one school needs to respond to external demands and what it has available (Cohen and Moffitt, 2009). Designers of accountability that want to reach their aims must address this gap and try not to exacerbate the gap between what a school must do and what it can do – else the accountability scheme will not realize the improvement it seeks.

Conventional test-based accountability has assumed two rational points to be true about the relationship between external and internal conditions: All internal environments will be able to respond, if provided the right incentives; and schools possess the resources they need to hit specific targets set by external policies. Based on these assumptions, the prevailing designs for school accountability have assumed that standard, one-size-fits-all approaches will lead to average improvements across schools.

While these assumptions hold for some schools, they do not actually hold for all schools –rather, they grossly underestimate diverse realities inside schools and the

resources available to schools to be mobilized towards meeting the external demands for improvement and sources of where and how schools acquire new resources. While some teachers possess the knowledge and skill they need to improve their teaching and some schools cultivate cultures in which expert knowledge and norms for problem-solving actually supersede any external mandate, many schools do not. Many schools possess internal environments – norms and standard operating procedures that offer few resources, or worse, conflict directly with the aims of external pressures. While the conventional models offer incentives for schools to acquire new resources and develop new practices, it is increasingly clear that incentives without some support for finding resources is insufficient (Massell and Goertz, 2005; Finnigan and Daly, 2012; O'Day, 2002).

The case of New York City's school accountability tools, reveals an effort to balance the external and internal environments through its development of two tools – The Progress Report, an external monitoring tool that was used to measure school's influence on students' learning and a second one, the Quality Review, which was used to make explicit which structures and professional norms that schools needed to attend to or develop in order to improve instruction and student learning. In addition, the designers sought to supply schools with supports in the form of networked communities that were to serve as resources to address capability gaps. Due to a bi-level accountability scheme that held networks accountable for school outcomes, everyone had strong incentives to pay attention to what was specified by the tools and make sense of them in relation to their unique internal circumstances.

The combination of the Progress Reports and Quality Reviews was particularly interesting because it combined a somewhat conventional method of telling all schools that they had to hit certain targets with an unconventional mechanism that specified the internal aspects schools needed to attend to in order raise student achievement. Both of these tools tried to encourage an external standard of quality that acknowledged diverse ways of achieving the standard – thus encouraging innovation, which Mayor Bloomberg and Chancellor Klein sought to create with new structures. One way they attempted to create structures that could do this was to create external evaluation measures that were standard, specific (so as to be instructive) and open.

For example, the Quality Review evaluated schools on whether pedagogy in a school was aligned to a coherent set of beliefs about how students learn best (Indicator 1.2). The central office did not tell schools what beliefs they needed to hold; rather, it was up to the schools to determine their coherent beliefs and how they would align pedagogy accordingly. This had the effect of focusing everyone's attention on particular aspects of internal environments that in the past had remained completely unspecified by accountability schemes – things like the relationship between coherent beliefs and pedagogy in a school. Never before in the U.S. has a central office said to schools, we are going to judge you on how well pedagogy across classrooms within a school is aligned to some visible set of shared beliefs. Instead, most accountability models left the "black box" of pedagogy untouched. The models told schools, we are going to judge you on how well your students do on standardized tests even though the tests may or may not relate to a curriculum that you teach (For example, many states use the ACT as the State's 'standardized test', but the ACT is an aptitude test – unrelated to standards for learning or curriculum). By contrast, the NYC approach sought to measure schools' internal conditions and left room for schools to be distinct in their choices. For example, one senior executive explained that they were trying to answer, "What is the best way to create a tool that standardizes quality, but accounts for diversity of performance and quality?" The combination of the tools was their attempt to answer that question, and because these were accountability tools, schools and networks were positioned to understand how that applied to each school's unique situation. This allowed schools and networks to use the accountability tools as guides for organizing their relationships and creating targeted supports matched to each school's unique needs.

New York City used a top-down approach to draw attention to professional, bottom-up changes in school's internal environments, in part, because they acknowledged that a handful of schools were well positioned to respond to the pressure for making progress with all students, but the district's culture was not. They used school accountability to motivate the creation of new mental models. For example, one architect explained,

> …in 2007, when I first went there, I would go to meetings with principals where they would say outright to each other in a way that was not controversial or

whatever, they would say just say, there are students that we just can't reach

because they can't do it – their families, their neighborhoods, their poverty – they

would say, we can't fire teachers so you can't expect us to do anything. So, we

just kind of do the best we can and that's mostly for some subset of students and

then there's always going to be some students who just aren't reachable.

And so, we talked about how you start, in a system like that, to get people to

adopt a sense of responsibility, and I think that's the real issue with the

Progress Reports that they were designed to do that – to create a sense of

external responsibility and then with the other mechanisms like inquiry teams and

networks and like that could transform into internal accountability that people

would feel.

When people feel responsible for the same things that are measured in evaluations, there is stronger alignment between the external and internal environments. Prior studies have shown that schools equipped with internal norms and shared conceptions of collective goals aligned to external environments tended to have stronger organizational capacity to adapt and improve student performance (Elmore et al, 2004; Newmann, King and Rigdon, 1997). Stronger organizational capacity tended to equate with stronger technical cultures and command over resources and development of capabilities that helped the teachers in these schools solve problems of practice and increase student achievement. Past accountability models did not address the alignment, but assumed that the right incentives would force schools to align themselves accordingly. New York City took a different stance and said that they would hold schools directly accountable conditions that would encourage better alignment.

Mayor Bloomberg and Chancellor Klein targeted accountability metrics and supports like the networks for using those metrics because they wanted to change the district's culture – from a culture of "excuses," "compliance," and "top-down bureaucracy" to a culture "where adults take responsibility for ensuring that all children regardless of their circumstances learn and achieve…and the central focus is on results and doing whatever it takes to help each student...individual great schools, where principals and their teams design the programs that their particular students need to succeed" (NYC DOE, 2008, pg. 4).   The Progress Reports and Quality Reviews were

used to evaluate schools, which had the effect of communicating a need to shift culture – across the system, including central office, networks and schools. However, for this messaging to become embedded in schools' cultures such that educators would adopt a sense of responsibility for educating and serving all students, NYC had to make important changes that affected the relations between internal and external conditions – beyond top-down evaluation tools.

One of the key changes in NYC was that the Chancellor granted principals the authority to make decisions about the internal conditions in their schools, in exchange for stricter accountability. By granting all principals authority (not just the high-performing ones as was done in other cities), New York City permitted the principals to have direct control over the factors that influenced their outcomes. This was one attempt to create the conditions (perhaps not the knowledge or skill) that made it possible for principals to foster internal motivation and commitments for improving teaching, solving problems related to instruction, and helping their lowest performing students – regardless of a school's past performances or the population of students served.

Yet, central office was aware that they could not just give principals freedom and set targets for them to hit; rather, they were aware that there were other conditions that central office had to put in place to help principals transition and promote the development of capabilities inside schools, and that structural changes they put in place would cause all manner of complications for school leaders. One of the original architects explained that the tools were meant to help initiate *and* support the transition among leaders in their schools. He explained,

> But I think it was really clear that we knew that we needed to motivate people to be attentive to how their students were doing and that that had to be focused on every single student and we wanted to instill a sense of deep responsibility for all of the kids. And secondly then, we wanted to create the resources and tools that would enable educators who came to that view and said, okay I feel responsible and I see failure happening, provide them with a set of tools that they could use in a strong way and we realized this would take a lot of work too, to provide those tools to them.

NYC's Department of Education used the accountability tools to motivate people to attend to all students learning experiences and shift thinking about what people responsible for, but they were also aware that they needed to provide "resources and tools" in order for people once they came to accept that they were responsible for all students and "use [the tools provided] in a strong way." Yet, the architects were very aware that this was a terribly ambitious assignment they had given themselves. They "realized this would take a lot of work, too, to provide those tools to them."

A second aspect meant to promote wider acceptance of central's messaging was the establishment of the networks as support organizations. Unlike other districts, where only low performing schools, or those targeted for interventions due to persistent underperformance, NYC attempted to create networked communities of schools that could foster learning and development of capabilities across and within schools. These organizations were specifically designed to be *learning* supports; they did not possess authority to evaluate or hold schools accountable, which created all manner of complications, but the intention was to help schools where they needed it most and differentiate support across schools such that engagement with improvement efforts was driven by internal actors, but supported by the networks that were considered to be external actors.

Because the network teams were not principals' rating officers (superintendents were rating officers), the networks could be intermediaries between central office and schools. The arrangement was meant to encourage engagement with reform principles and prompt commitments from principals and their teams to develop the capabilities they needed to improve instruction (indicator 1.2); build coherence across classrooms tied to standards (indicator 1.1); maintain a culture of trust (indicator 1.4) and high expectations for learning among staff and students (indicator 3.4); and to establish routines for structured inquiry focused on improving learning (indicator 4.2) and monitoring and revising systems (indicator 5.1).

The designs for the Quality Reviews and Progress Reports were intentional attempts to focus attention and guide behavior and direct the development of capabilities towards particular goals; the networks were designed as supports to encourage participation and negotiation of meaning around the tools within the boundaries of each

school's unique internal conditions. As one architect explained, "we were very focused on both sides of that issue, accountability and chief resources." Similar to the way that Stein and Coburn (2008) describe the "architectures for learning" that districts in their study designed to help teachers implement a new math curriculum, the architects in NYC were trying to "…create two kinds of affordances for learning: (a) getting boundary objects in place so that practice at various levels has to be organized around them; and (b) getting the right people to interact to make something happen" (pg. 617). In NYC, they were trying to strike the right balance between these two affordances.

What we learn from NYC's attempts is that finding that balance is about attending to learning – the learning that people must do to use the tools in the unique circumstances and then learning from those opportunities about how to adjust or adapt tools or interactions – none of these things can be known at the start of an initiative. This means that designers should not presume to have perfect solutions from the start; rather, they must remain open to learning and asking questions. For example, during a conversation with one of the chief architects, he reflected on what they were trying to accomplish with the designs for the Progress Reports and Quality Reviews. He explained,

> For me, the hardest part of what we were trying to do was to get that balance
> between the internal and the external. The external was there in my mind to drive
> and support and create some structure around the internal but it was not, by any
> means, to replace the internal or to deny that the internal existed; it was all in
> service of that...then the question is just how you get to that right balance."

The lesson is that there is not a recipe, or answer that tells districts how to achieve perfect balance; rather, the lesson to take away is that, perhaps the best districts can do, is try to strike a balance between external and internal and learn from their efforts.

Bold and incremental changes

The design and implementation of the accountability tools in NYC were bold departures from earlier efforts to manage school improvement in New York City. In the past, the thirty-two districts managed reform efforts separately, but this led to vastly inequitable educational experiences for students dictated mostly by geography - whether a student attended school on the Upper East Side of Manhattan (District 2) or some place else like

the South Bronx (District 7); a student's zip code was a strong predictor for educational outcomes.

When Mayor Bloomberg was awarded control of the city's schools, he sought to radically overhaul the system to make education quality a matter of school performance instead of neighborhood. He drastically changed the governance structure, and in doing so, the relationship between central office and the city's schools.

The strategic changes that the Mayor and Chancellor Klein made were meant to be bold in order to achieve bold improvements. A report issued by the Chancellor's office explained that they sought to create stability and coherence in the system by setting new academic standards, implementing a core curriculum, building a new, streamlined management structure, creating a new leadership development program and new accountability tools to "hold everyone in the system accountable for results." The report explained, "Once we created stability and coherence in the system, we took the logical next steps by focusing on the three areas that are most likely to produce the bold improvements our system needs and our children deserve: leadership, empowerment, and accountability" (NYC DOE, 2008, pg. 3).

The point of many of the Children First reforms was to disrupt old ways of working; the leadership took bold action and made radical decisions. However, they risked confusing everyone and distracting people with the next new thing, if they jumbled the roll out of changes, which would have made it less likely that the reforms would have been able to trigger deep, sustainable changes in practices. Yet, if they went too slow, they would not accomplish what they set out to accomplish within the Mayor's time in office. They had to balance bold disruptions with learning so that the reforms would be taken up by principals, teachers, networks, superintendents and central office – everyone in the system needed to be on board with the new reforms. To manage this tension, the DOE created communications structures – including the networks to help turnkey many of the changes that were made throughout the implementation of reforms. In addition, the DOE created new roles including Parent Coordinators who were full time employees in every school to be able to help parents understand the changes.

In addition, the architects of the accountability system in NYC understood that designs matter a great deal and that they can evolve. As a result, they specifically built

the designs to do just that and created feedback loops and developed internal capabilities within the central office to manage the process of learning from and through implementation. They used continuous improvement practices to manage the improvement of the designs to fit the strategy for reform. For example, one architect explained, "We applied a philosophy of continuous improvement to work we were doing…we would do it and we would manage the change. The change management aspects it were a challenge because schools would complain that it was a moving target for their work. " To address this, the DOE created a "reflection period" in which the DOE would conduct more than seventy sessions twice a year to talk with networks and principals and allow social opportunities for people to negotiate and make sense of the tools at whatever point in the evolutionary cycle they may have been. The strategic use of "change management" techniques by the central office helped manage the tensions between incremental and bold change.

What we learn from this story of changing accountability in NYC is that significant changes in structures can happen relatively quickly, but it required taking the system as the unit of change. The story in NYC under Mayor Bloomberg is in line with McLaughlin's and Talbert's (2003) study of urban districts in California. These authors found that reforming districts had common characteristics to their reforms, including, that all of the districts took a system-level approach to reform, sought to create learning communities within central office (not just schools), established coherent focus on teaching and learning, and instituted data-based inquiry and accountability. The authors characterized reforms as weak and attributed it to weak coordination at the central office and strong, if the central office coordinated learning throughout the system. In NYC, they definitely took the approach that they were trying to change the system and sought to reconstruct central office in the process. Part of the reconstruction was to use central office as a facilitator of learning – responsible for managing the conditions that would make learning about and from the reforms possible while still moving the entire system toward the visible goal of making all schools valuable places for all students.

Summary

In many ways, the tensions and emerging themes from the story of New York City's school accountability reforms that I described in this dissertation are reminiscent of the cleavages in the early standards-based reform movement in the 1990s. Drawing on quotations from *Education Week* articles from 1993, Mehta (2013) highlights that the standards-based "reform strategies can be roughly divided into two camps: those who believe that schools must change from the ground-up, one building at a time; and those who believe that the federal and state policies should provide the stimulus and the conditions for building-level change" (pg. 214). Those who thought change would come from the ground-up, one building at a time, saw teacher capabilities as the most important factor contributing to weak schools. Mehta refers to this as the "soft" view that believed that what was needed to change schools was greater emphasis on helping teachers develop skills and knowledge. This group saw the need for standards to serve as instructive guidelines from which professional development could follow.

By contrast, the "hard view" saw the most pressing problem to be that the structures around schools breed complacency. Those who took the hard view argued that change would result from motivating people to change – hold them accountable for results. Mehta (2013) suggested that the combination of these two views would eventually become the template for standards based reform in the U.S. In 2002, the passage of No Child Left Behind signaled that the hard view should be the favored view in order to reduce achievement gaps and make sure that every child that attends public schools is proficient in reading and writing. Yet, Mayor Bloomberg took a markedly different approach.

In New York City, they tried to strike a balance between hard and soft approaches to changing student achievement by emphasizing results and holding everyone in the system accountable for improving student achievement, as well as, focusing attention on the development of continuous improvement capabilities within each school at the ground level. They attempted to do this at a scale and pace unrivaled in education reform. The scale of it suggests that we could export their approach and see rapid results across the country. However, it is unlikely that the reforms could be replicated exactly;

there were certain factors involved in NYC's efforts that were fairly atypical in education reform.

Mayor Bloomberg was able to secure significant support from private philanthropy and business to help incubate some of the initiatives. For example, a private contribution of $75 million was used to launch the New York Leadership Academy that was created during the beginning phases of the reforms create a talent pipeline of new leaders equipped to lead in a new kind of system. The Leadership Academy was modeled after General Electric's Crotonville facility for managers. Jack Welch, the former CEO of General Electric, who was credited with saving General Electric was one of the Academy's board members (Goodnough, 2003; Bloomberg, 2003). Similarly, money from private donations was used to support the development of initial strategy. When Joel Klein was first named chancellor in 2003, he underwent a three month planning process that was funded with $4 million from The Broad Foundation and others (Gyurko and Henig, 2010).

Mayor Bloomberg and Chancellor Klein pursued philanthropic dollars fairly aggressively and relaunched the Fund for New York City Public Schools (Kelleher, 2014). This fund helped foundations like the Bill and Melinda Gates Foundation directly invest in particular projects. The Bill and Melinda Gates Foundation invested more than $150 million dollars in New York City's high school reform and small schools efforts.

While funds from philanthropy accounted for less than 0. 5 percent of the system's budget, these investments provided for innovation and helped build needed infrastructure like the ARIS data management system that would serve as the backbone of the accountability tools (Kelleher, 2014). One former executive explained, "…there was never a question, 'do we have the money to pull this off?' There was never a question about typical resources like money. We had whatever we needed. If we needed to designers to come in and build us a tool, we got it. If we needed a company from England to help build the Quality Review for 4, 5, 6 billion, then you got it. There were no resource constraints." These investments and the fact that "Children First was led and staffed by a small army of private consultants and educators new to or from the periphery of the New York City school system" (Gyurko and Henig, 2010, pg. 95) helped contribute to a culture that was described to me by one architect as "mission-driven start-

up." Few educational bureaucracies around the country have access to unlimited resources or are able to cultivate cultures that seem similar to that of a "mission-driven start-up." Rather, most districts must contend with financial and political constraints like answering to interest groups and constituents that NYC DOE was excused from under Mayor Bloomberg's unconventional political leadership.

When Mayor Bloomberg took over control of the schools in 2002, less than half of the students in New York City graduated with a Regents Diploma and less than forty percent of students in grades three through eighth were proficient in math and reading (NYC DOE, 2008). When he left office in 2013, high school graduation rates had improved by nearly twenty percentage points (NYC DOE, 2013) and one of the most positive outcomes was a reduction in the racial disparity of who graduates (Kelleher, 2014). More students in all grades were proficient in math and reading. There is strong evidence that Mayor Bloomberg's reforms improved student outcomes.

Yet, there were so many substantive systemic reforms put in motion over the Mayor's eleven years in office, including school finance, governance, human-capital pipelines, school structure, and accountability that "researchers caution that the complexity of interlocking reforms makes it difficult to determine what strategies and/or interplay among strategies produced which results" (Kelleher, 2014, pg. 10). Kelleher quotes Kemple saying, "It's like mercury – as soon as you put your finger on it, it shifts." However, perhaps that is the point – that system level reform should be less about isolating impact and more about seeing the complexity and interplay of the environment as a set of conditions for reform (Peurach, 2011). There is no doubt that NYC's accountability scheme was concerned with changing outcomes and creating a results-oriented focus, but it also tried to strike a balance by focusing on the conditions for achieving different results by using accountability to do both.

References

Bloomberg Magazine (2003). Can Bloomberg Fix New York's Schools?
Retrieved: www.bloomberg.com/bw/stories/2003-06-08/

Cohen, D.K., Moffit, S.L. (2009). *The ordeal of equality: Did federal regulation fix the schools?*. Harvard Education Press. Cambridge, MA.

Goodenough, A. (2003). Executive who saved G.E. is to train school principals. *The New York Times.* January 14, 2003.

Gyurko, J., Henig, J.R. (2010). Strong Vision, learning by doing or the politics of muddling through? In Katrina Buckley, Jeffrey Henig, Henry Levin (eds). *Between Public and Private: Politics, Governance, and the New Portfolio Models for Urban School Reform*. Harvard Education Press: Cambridge, MA.

Kelleher, M. (2014). New York City's Children First: Lessons in school reform. Center for American Progress. January 2014.

Klein, J. (2014). *Lessons of Hope: How to fix our schools.* Harper Collins: New York, NY.

McLaughlin, M.W., Talbert, J.W. (2003). Reforming Districts: How Districts Support School Reform. Center for the Study of Teaching and Policy University of Washington.

Mehta, J. (2013). The Allure of Order: High hopes, dashed expectations, and the troubled quest to remake American schooling. Oxford University Press.

New York City Department of Education (2008). Children First: A Bold, Common Sense Plan to create great schools for all New York City children. 2008-09.

New York City Department of Education (2013) What's Next for School Accountability in New York City? Released November 2013.

Parthenon (2013). An Assessment of the New York City Department of Education School Support Structure. Report November 2013.

Peurach, D. (2011). Seeing complexity in public education: Problems, possibilities and success for all. Oxford University Press.

Stein, M.K., Coburn, C.E. (2008). Architectures for Learning: A Comparative Analysis of Two Urban School Districts. *American Journal of Education.* Vol. 114, No. 4, pg. 583-626.

Wohlstetter, P., Smith, J., Gallagher, A. (2013). New York City's Children First Networks: Turning accountability on its head. *Journal of Educational Administration.* Vol. 51, no. 4, pg. 528-549.

CHAPTER VI

EPILOGUE

In November, 2013, Bill de Blasio won the mayoral election in New York City, which gave him control of the city's public school system. As mayor, Bill de Blasio began to dismantle many of the reforms that were put in place under his predecessor, Michael Bloomberg.   One of the first things to be eliminated was school closure as a response to school failure.

During Mayor Bloomberg's tenure, the Department of Education shut down or phased out more than 157 schools due to persistently low scores on the Progress Reports and Quality Reviews. Although, the Bloomberg administration was also responsible for opening 656 new, smaller schools – of which 176 were charter schools (Taylor, 2015), the closures garnered more public attention – much of it criticism.

Unlike Mayor Bloomberg, who courted successful charters to join the portfolio district, Mayor de Blasio waged a war with charter schools and openly opposed opening any new ones under his watch. As evidenced by his very public fight with Eva Moskowitz, a Democrat like the Mayor and a former city council member (chairwoman of the Education Committee), the problem was not politics. Moskowitz, founder of Success Academy Charter Schools, was a fierce advocate for charters; her network performs in the top 1 percent of all schools in the state (Bergner, 2014). In Success Academy, a school located in Bedford-Stuyvesant, a low-income neighborhood with a reputation for violence and crime, where 95 percent of the students are black or Latino and most qualify for federally subsidized lunch, 98 percent of Success students scored at or above grade level in math (80 percent received the highest ratings). This school is not an anomaly among Success Academies, but it is citywide. In NYC, "fewer than one-fifth of black students in the city can read or do basic math at grade level" (Bergner, 2014). Given Success Academy's achievements, the Mayor's animosity towards Moskowitz and Success Academy Charter Schools appeared perplexing.

Mayor de Blasio opposed charters suggesting that they take resources away from public schools - "robbing the many to teach the few" (Bergner, 2014). This moral and openly adversarial stance against charters, even those that have proven that they raise student achievement, was antithetical to his predecessor's approach to fixing the schools. Mayor de Blasio favored keeping traditional public schools open and promised to provide struggling schools with new "supports."

In January, 2014, Mayor de Blasio announced that Carmen Farina, a former educator, would serve as Chancellor. The Mayor and Chancellor eliminated the A-F rating system for schools that had been a controversial part of the Bloomberg administration's accountability system. Under Chancellor Farina's leadership, the Department of Education instituted School Quality Reports to replace the Progress Reports and Quality Reviews. There are two kinds of School Quality Reports – a snapshot for families and a guide for educators. While the new reports rate schools on all of the same metrics used by the Bloomberg administration, including the Quality Review, student progress, student achievement, school environment, college and career readiness and closing the achievement gap, the new reports look different and use four color-coded levels to designate school rankings: Exceeding target, Meeting Target, Approaching Target and Not Meeting Target (NYC DOE, 2015). The snapshots that are for families give schools a rating of Excellent, Good, Fair or Poor in the different areas.

In January 2015, Chancellor Farina announced that the network support structure would no longer exist. She reinstated superintendents as the primary source of operational and instructional support for schools, which would be located in Field Borough Offices and supervised by superintendents. Under this arrangement, principals would no longer select their support provider. Instead support returned to its former form and would be a geographically-based bureaucracy.

Ms. Farina explained that she wanted to revert to the formal hierarchy that existed before Mayor Bloomberg because it would offer clear lines of authority. According to reports in the *Wall Street Journal*, the Chancellor shared that having the superintendents in charge would be mean that, "They will be my eyes and ears; they answer to me" (Brody, 2015). However, it was not clear whether superintendents would be held accountable for their schools' outcomes.

On the last day of their session, the legislature in Albany passed a law to renew mayoral control of New York City schools for one year, which was "far from the permanent extension sought by Mayor de Blasio" (Kaplan, 2015). There were reports that the state legislature refused to extend mayoral control for longer because they were critical of Mayor de Blasio's and Chancellor Farina's School Renewal program, which was the program that Chancellor Farina designed and implemented to target supports for the city's lowest performing schools (Decker, 2015).

The turbulent times for the current administration in NYC highlight how challenging education reform can be – especially as Albert Shanker observed so many years before: "The key is that unless there is accountability, we will never get the right system…as long as the discussion is not about education and student outcomes, then we're playing a game as to who has the power." Whether the mayor wins out and keeps his power over the schools, or control shifts back to thirty-two regions, or some other governance structure emerges, it is clear that New York City will remain an interesting subject from which we can learn about system reform, school accountability and their intersections.

References

Bergner, D. (2014). The Battle for New York Schools: Eva Moskowitz vs. Bill de Blasio. *The New York Times Magazine.* September 3, 2014. Retrieved 6/29/15: http://www.nytimes.com/2014/09/07/magazine/the-battle-for-new-york-schools-eva-moskowitz-vs-mayor-bill-de-blasio.html

Brody, L. (2015). New York City Schools Chancellor Boosts Superintendents' Clout. *Wall Street Journal.* January, 23, 2015. Retrieved 6/18/15: http://www.wsj.com/articles/new-york-city-schools-chancellor-boosts-superintendents-clout-1422023365/

Decker, G. (2014). Senate leader: Questions about NYC schools spending fueled one year mayoral control deal. *Chalkbeat: New York.org.* Retrieved 6/29/15: http://ny.chalkbeat.org/2015/06/25/senate-leader-says-questions-about-nyc-ed-spending-fueled-one-year-mayoral-control-deal/#.VZHjb6bptg4

Kaplan, T. (2015). Wide-ranging deal approved as Albany session closes. *The New York Times.* June 25, 2015. Retrieved 6/29/15: http://www.nytimes.com/2015/06/26/nyregion/new-york-legislative-session.html

New York City Department of Education (2015). High School Quality Guide: Educator Guide. Retrieved: http://schools.nyc.gov/NR/rdonlyres/F594D93F-393D-4D67-A5F6-EB1076B1CF94/0/EducatorGuideHS1202015.pdf

Simon, S. (2013) . New York City Democrats embrace full speed reversal on education reforms. *Politico.* Retrieved: http://www.politico.com/story/2013/09/new-york-city-democrats-education-reform-96632.html

Taylor, K. (2015). Chancellor Carmen Farina Changes New York City Schools' Course. *The New York Times.* February 6, 2015. Retrieved 6/18/15: http://www.nytimes.com/2015/02/08/nyregion/chancellor-carmen-farina-changes-new-york-city-schools-course.html?_r=0