School Principals under Accountability

By

Hajime Mitani

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

August, 2016

Nashville, TN

Approved:

Jason A. Grissom, Ph.D.

Gary T. Henry, Ph.D.

Ron Zimmer, Ph.D.

Mark Lipsey, Ph.D.

ACKNOWLEDGEMENTS

I would like to give thanks to all of the individuals who have supported me in this dissertation process with patience and encouragement. First, I would like to thank my wife, Mayumi Mitani, for her support in many aspects during the five years of my graduate work. She provided me with love and spiritual support, which allowed me to complete my dissertation research. I thank my three children, Maki, Mao, and Hozana Mitani, for giving me joys and happiness as their father and allowing me to work on this dissertation even during the weekends. I am also grateful for both of my parents, Michiki and Kaoru Mitani, and both of Mayumi's parents, Tamotsu and Kumiko Sugiura. They encouraged me to complete this important step in my academic career and gave my wife and children a plenty of gifts. I thank Mayumi's friends, especially, Reina Lyons. She helped Mayumi raise three kids when I was busy with my work. Spiritual support provided by a former missionary, Sue Benedict, and Pastors Stephen Young, Charlie Williams, and Neal Hick was invaluable to me and my family members.

I also would like to give thanks to all of my dissertation committee members – Jason Grissom, Gary Henry, Ron Zimmer, and Mark Lispey. First, I am very thankful to my academic advisor, Jason, for his continued support during the five years. He gave me significant research training that enabled me to grow as an educational researcher and start my academic career as an assistant professor after graduation. I was fortunate to be able to work with him for the entire five years. I thank Professor Gary Henry and Dr. Mark Lipsey for giving me insightful comments to polish my dissertation work, especially methodologies that I chose to use. Without their technical support, I would not have been able to complete this dissertation. I am grateful that Professor Ron Zimmer gave me various supports. First, as the director of the Graduate Study, he helped me secure fifth year funding and summer

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure   Page

Chapter 3

Chapter 4

CHAPTER 1

INTRODUCTION

Objectives

The federal No Child Left Behind (NCLB) Act of 2001 introduced a federal-level

performance-based school accountability system in the U.S. public education system. This

federal law symbolizes the school accountability movement since the 1990s. The major goals of

the law were to raise student performance in math and reading, and narrow achievement gaps

among different student subgroups (U.S. Department of Education, 2001). This system punished

low-performing schools that continuously missed performance targets set by each state education

agency through sanctions such as school choice options, staff replacement, school closure, and

school restructuring.

The assumption behind this system was that the sanction system would provide adequate

incentives for educators and administrators to change their behavior patterns and daily practices

ultimately resulting in improvement in student test scores in math and reading. A number of

studies have investigated this assumption by examining the effects of NCLB on students as well

as teachers, both qualitatively and quantitatively. Some of these studies generally find that

NCLB improved student performance, especially in math, and other academic outcomes (e.g.,

Dee & Jacob, 2011; Lauren & Gaddis, 2012; Reback, Rockoff, & Schwartz, 2014). Other studies

suggest that, while NCLB changed the teachers' perceptions of cooperation with their colleagues

negatively, it did not change teachers' job satisfaction or commitment level (Grissom,

Nicholson-Crotty, & Harrington, 2014). Researchers, however, also report that teachers viewed

the NCLB system negatively and that it increased the amount of pressure and job stress among teachers, which led some teachers to burn out or become demoralized (e.g., Berryhill, Linney, & Fromewick, 2009; Center on Education Policy, 2006; Dee, Jacob, & Schwartz, 2013; Mertler, 2011; Santoro, 2011; Sunderman, Tracey, Kim, & Orfield, 2004).

Despite the large number of studies on the effects of NCLB on students and teachers, few studies have investigated its effects on school principals. While principals' influence on student achievement may be indirect, they play an important role in improving school and student outcomes (e.g., Branch, Hanushek, & Rivkin, 2012; Brewer, 1993; Grissom, Kalogrides, & Loeb, 2015; Grissom, Loeb, & Master, 2013; Hallinger & Heck, 1998; Waters, Marzano, & McNulty, 2003). One report even claims that school leadership is "second only to classroom instruction among school-related factors that affect student learning in school" (Wallace Foundation, 2013, p. 5). Some of the principals' roles include establishing academic curriculum, observing and evaluating teachers, hiring and retaining teachers, building professional learning communities, maintaining school facilities, working with parents and local community organizations, communicating with district and state administrators, among others (e.g., Leithwood, Louis, Anderson, & Wahlstrom, 2004; Wallace Foundation, 2013; Waters et al., 2003).

Given these important roles, it would be reasonable to think that NCLB might have influenced school principals some ways. In fact, school principals were the prime target of NCLB or, more specifically, the NCLB sanction system. They had to make a lot of important decisions in order to improve school and student outcomes and meet performance targets set by the state education agencies. When their schools continued to perform below the set goals, they became a target of replacement. In addition, principals had to comply with NCLB's numerous administrative requirements such as data reporting (Koyama, 2014). This system, therefore,

made the already complicated job even more complicated and challenging (Brown, 2006), which could have increased the principals' job stress. Indeed, survey studies find that a majority of principals had unfavorable views toward the NCLB system (Educational Testing Service, 2008; Lyons & Algozzine, 2006; Salazar, 2007). Increased job stress could lead principals to burn out and eventually leave their schools, especially when they faced NCLB sanctions. Recent studies find that NCLB increased principal turnover rates (Ahn & Vigdor, 2014; Li, 2012).

Despite these compelling anecdotes, few studies have systematically explored the extent to which NCLB influenced school principals. This dissertation fills this gap in the literature and focuses on to what extent the NCLB sanction system is associated with school principals' working conditions, job satisfaction and stress, and turnover behaviors using longitudinal administrative data in the state of Missouri and responses from a national representative sample of school principals to the Schools and Staffing Survey (SASS) administered during 2007–08 by the National Center for Education Statistics, the U.S. Department of Education (USDOE).

Poorly designed incentive programs often backfire because either their incentive structures, performance measures, or reward criteria do not capture true values of workers' performance or do not align well with their professional or personal standards (e.g., Dixit, 2002; National Research Council, 2011; Podgursky & Springer, 2007). If the design of the NCLB sanction system was inappropriate, the system might negatively impact principals' working conditions and increase their job stress. Poor working conditions could lead to turnover among principals. While turnover may be a good outcome for schools if ineffective principals systematically leave their schools, it could become a problem when it happens frequently, irrespective of whether principals are effective or ineffective. Instability in school leadership often causes conflict between teachers and new principals, disruption of the instructional

3

program, and changes in school visions, all of which likely prevent schools from improving student outcomes (Hargreaves & Fink, 2006; Meyer, Macmillan, & Northfield, 2009). This organizational turmoil could negatively impact student performance, as recent studies find (Béteille, Kalogrides, & Loeb, 2012; Miller, 2013). It is, therefore, meaningful to examine whether NCLB sanctions are associated with principals' working conditions, job stress, and turnover behaviors.

Overview of Studies

The purpose of this work is to explore to what extent the NCLB sanction system influenced school principals. It examines the design of the sanction system and its influence on principals' working conditions, job stress, and turnover behaviors. This dissertation is comprised of three studies. The first study broadly reviews prior studies of NCLB, discusses theories of incentives and motivations behind the NCLB sanction system, and uses them to assess the design of the NCLB sanction system. The second study empirically examines to what extent NCLB sanctions are associated with principal turnover behaviors using longitudinal administrative data and detailed school-level assessment and AYP (Adequate Yearly Progress) data obtained from the Missouri Department of Elementary and Secondary Education. The third study turns to the national level analysis and investigates whether NCLB sanctions are correlated with principals' working conditions, job stress, and turnover behaviors using the Schools and Staffing Survey 2007–08 and detailed school-level assessment and AYP data systematically obtained from 45 states. The following subsections describe each of the three studies in detail.

Study 1: Influence of NCLB: A Review of Literature, Theories behind the NCLB Accountability System, and An Examination of the System

In this study, I first review prior studies that examine the effects of NCLB on students, teachers, and school principals. As one of the key features of NCLB was the performance-based sanction system, I also review studies that investigate the effect of each sanction. NCLB sanctions included district technical assistance, school choice options, supplemental education services, corrective actions, and school restructuring.

Following the review, I discuss theories of motivations and incentives, on which the NCLB sanction system was based. Policymakers assumed that the NCLB sanctions would provide school principals adequate incentives to change their behaviors such that their schools meet the performance targets set by each state education agency. However, if the design of the sanction system was inappropriate, school principals might not respond in the expected ways. I first define motivations and incentives and later discuss under what conditions a person responds to incentives based on studies in psychology, neuroscience, and behavioral economics. I also discuss the "undermining effect," which means that a person's intrinsic motivation reduces once an incentive is removed. The current literature is divided on this topic and it is important to discuss it.

After the theoretical discussion, I identify common problems inherent in incentive systems. For example, if a performance measure used in an incentive system does not capture true values of a worker's productivity, incentive may not work. This is often a problem when a job has multitasking features. The teaching profession and the principalship are typical jobs with such features. Another example is performance target. If a performance target is unreasonably high or low, a person currently performing far above or below the target may not respond to the incentive.

Based on these common problems, I discuss the appropriate design of an incentive policy. I provide a general framework to assess whether the design of an incentive policy is appropriate to incentivize people and achieve the desired goals. Then, I use this framework to assess the design of the NCLB sanction system.

In the last section, I discuss labor market implications for school principals. For example, if the design of the NCLB sanction system was not appropriate, principals might have greater job stress, which could lead them to transfer to different schools, change their positions, or exit the public education system. I discuss these possibilities.

Study 2: Principal Turnover under NCLB Accountability Pressure

In this study, I empirically estimate the relationship between NCLB sanctions and principal turnover. First, I focus on the first-time failure. This means that a school missed adequate yearly progress or AYP for the first time and thus faced an informal sanction for the first time. Informal sanction means that schools that missed AYP were exposed to public criticism and scrutiny as the state education agencies released school report cards to the public every year. I differentiate the first-time failure from other times because it could have been a more shocking event to principals. It might be the time when principals started feeling the accountability pressure. I answer the following research questions:

(1-a) To what extent is the first time informal sanction associated with principal turnover?

(1-b) Is this association different by turnover types (i.e., transfer, position change, and exit)?

(1-c) Do principal qualifications and/or school characteristics moderate this association?

The third research question comes from an idea that principals with different qualifications may have different skills to cope with accountability pressure. For example, experienced principals might possess better knowledge on education production functions to improve school performance than inexperienced principals. Similarly, principals serving certain groups of students might face different and harder challenges when they faced the informal sanction.

Next, I shift the focus from the first-time informal sanction to all informal sanctions and the formal NCLB sanctions. All informal sanctions include the first-time informal sanction and the non-first-time informal sanction. I answer the same set of research questions:

(2-a) To what extent are the informal sanctions and the formal NCLB sanctions associated with principal turnover?

(2-b) Is the association different by turnover types?

(2-c) Is the association moderated by principal qualifications and school characteristics?

For the formal NCLB sanctions, I focus on School Improvement Year 1 (SIY 1), which required districts to provide school choice options for eligible students. Title I school principals faced SIY 1 when their schools missed AYP for two consecutive years in the same subject (e.g., math and reading) or the same academic indicator (e.g., attendance rates and graduate rates). For the other formal sanctions (e.g., SIY 2 and SIY 3), I descriptively analyze the association, because a smaller number of schools actually faced these sanctions, which does not allow me to estimate the relationship precisely. Only Title I schools were subject to the actual NCLB sanctions, for SIY 1 and beyond, and therefore my samples are limited to Title I school principals only.

Finally, I examine patterns in principals' transfer and position change behaviors. For this analysis, I focus on the informal sanction as I observe enough principals who transferred and changed their positions. More specifically, I examine the following research questions:

(3-a) To what extent are characteristics of schools to which principals are transferred to differ from those where the principals used to work, and do changes in the characteristics vary by the informal sanction status?

(3-b) Are patterns in position changes associated with the informal sanction status?

To answer these questions, this study uses longitudinal administrative data and detailed school-level assessment and AYP data obtained from Missouri Department of Elementary and Secondary Education (DESE). The former data are available from 1993–94 to 2009–10, and I mainly use the data from 2001–02 to 2009–10. The school-level assessment and AYP data are available from 2004–05 to 2009–10. For years from 2001–02 to 2003–04, I utilize similar school-level assessment data maintained by the National Longitudinal School-level State Assessment Score Database and the Barnard/Columbia No Child Left Behind Database (Reback, Rockoff, Schwartz, & Davidson, 2011). I merge these data with NCLB sanction information collected from Missouri DESE and the U.S. Department of Education.

The primary methodological approach used in this study constructs a variable that measures the distance to the AYP threshold across subjects and student subgroups. It uses the distance variable as a key matching variable along with other principal, school, and district characteristic variables in the first-stage matching model and matches principals based on the probability of missing AYP. After the propensity score matching, I use a series of logistic and multinomial logistic regression models to control for the remaining differences in these characteristics and estimate the association between NCLB sanctions and principal turnover.

Study 3: Principal Behaviors under NCLB Accountability Pressure

In this study, I turn to a national sample of school principals and examine to what extent NCLB sanctions influenced principals' working conditions, job stress, and turnover behaviors. For NCLB sanctions, I focus on and combine the informal sanction and School Improvement Year 1 (SIY 1) together. The informal sanction includes both the first-time and non-first-time informal sanctions. I combine them because my data do not allow me to identify which schools faced the informal sanction in 2006–07. I answer the following two sets of research questions:

(1-a) Are NCLB sanctions associated with principals' working conditions and job stress?

(1-b) Is the association moderated by principal qualifications and school characteristics?

(2-a) Are NCLB sanctions associated with principal turnover?

(2-b) Is the association different by turnover types?

(2-c) Is the association moderated by principal qualifications and school characteristics?

This study relies on two key data sources. The primary data source on this study is the Schools and Staffing Survey (SASS) 2007–08 and the Principal Follow-up Survey (PFS) 2008–09. SASS is a survey of a national representative sample of teachers, principals, schools, library media centers, and districts that asked a variety of questions related to such topics as working conditions, instructional practices, and staff demographic and professional characteristics. I merge the SASS data with the PFS data, which followed up the sampled schools in 2008–09 to ask about their principals. I combine these data with the school-level Common Core of Data for 2006–07 maintained by the National Center for Education Statistics to supplement school characteristics data.

The other key data source is school-level assessment and AYP data that I systematically collected from 45 states. These data include information about the percent of students performing

at or above the proficient level in 2006–07 in each subject by student subgroups. I also collected

NCLB sanction information in 2006–07 from state education agencies and the U.S. Department

of Education.

The methodological approach used in this study to answer these questions is very similar

to the one used in the second study. I construct a variable that measures a distance to the AYP

threshold and use it as a key matching variable along with other principal, school, and district

characteristic variables in the first-stage matching model. After matching schools based on the

probability of missing AYP or the propensity scores, I use regression, logistic regression, and

multinomial logistic regression models to adjust for remaining differences and estimate the

association between the NCLB sanction and the outcome variables.

Contributions to Practice

This dissertation seeks to inform about our understanding of the influence of the NCLB

sanction system on school principals. Results from this research may provide useful implications

for policymakers. First, the first study offers a general analytical framework to assess the design

of the incentive policies. Many of the current education policies and programs include some kind

of incentives to motivate students, educators, and administrators. The framework would allow

policymakers to assess whether the current design of incentives is appropriate and most effective

to achieve their goals. They might be able to identify where the problems lie and adjust the

design appropriately to maximize the effectiveness of their incentive policies.

The second and third studies provide new empirical evidence on to what extent NCLB

sanctions influenced principals' behaviors. A lack of general understanding of how school

principals respond to accountability pressure would prevent policymakers from designing

appropriate accountability systems and incentive policies in general. Without knowledge, any of the accountability systems and incentive policies would have a potential to backfire and cause unintended consequences. These two empirical studies provide useful information for policymakers to improve the effectiveness of the current accountability systems and incentive policies that target school administrators.

Limitations

The empirical studies presented in this dissertation have several limitations. First, both studies use a propensity score matching method to reduce potential bias that comes from selection of schools into the NCLB sanctions. It matches principals based on observable characteristics. This method is based on the assumption that the assignment to the treatment (i.e., sanctions) is independent of the outcomes, conditional on characteristics of principals, schools, and districts. If this assumption is met, my estimates can be interpreted as the causal impacts of the NCLB sanctions. However, if this assumption is violated because of unobservable factors correlated with both the treatment status and the outcome variables, my estimates would be biased.

Secondly, these studies focus on the short term influence of NCLB sanctions on principals' working conditions, job stress, and turnover behaviors. Yet, NCLB sanctions could be more influential in the long run. Principals' job stress level in three years might be quite different from the level right after facing the sanctions. Thirdly, the third study uses the total number of hours that principals had spent on school activities and interactions with students during the school day, before and after school, and during the weekends as proxies for principals' working conditions. Yet, these two measures do not necessarily reflect the whole aspect of the working

conditions. As a result, findings on principals' working conditions need to be interpreted with caution.

Fourthly, both empirical studies investigate the association between the NCLB sanctions and principal turnover using reduced-form models, which do not distinguish voluntary turnover from involuntary turnover. If principal turnover results from district administrators' strategic human resource decisions rather than NCLB sanctions, then my estimates may not necessarily reflect the true relationship. In the first empirical study of principals in Missouri, I offer descriptive analysis of whether turnover was initiated by principals or district administrators by examining changes in principal characteristics before and after turnover.

Finally, my estimates may be subject to anticipation effects (see Malani & Reif, 2010). Principals might respond to NCLB sanctions before they actually faced them. That is, based on their schools' current performance, they might anticipate that they would face the sanctions in the near future so that they took actions before they were imposed with the sanctions. If this is the case, my estimates would be biased because those not facing the sanctions also responded to them. I argue, however, that, because these studies focus on early stages of NCLB sanctions, which were not severe sanctions, the last two problems may not necessarily threaten the validity of my results.

General Overview

The rest of this dissertation is organized as follows. In Chapter II, I review past studies that investigate the impact of NCLB overall and NCLB sanctions on (1) students, (2) teachers, and (3) principals. Then, I discuss theories of motivations and incentives, on which the NCLB sanction system was based. I discuss common problems inherent in incentive programs and offer

a framework to examine the design of an incentive policy. I use this framework to assess the design of the NCLB sanction system and then provide labor market implications for school principals that result from the design. Chapter III empirically estimates the association between the NCLB sanctions and principal turnover using longitudinal administrative data and detailed school-level assessment and AYP data from the Missouri Department of Elementary and Secondary Education. Chapter IV utilizes national representative data on school principals merged with detailed school-level assessment and AYP data to explore the relationship between the NCLB sanctions and principals' working conditions, job stress, and turnover behaviors. In Chapter V, I summarize findings from these studies and discuss the contributions to practice, policy implications, and suggestions for future research.

References

Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina*. NBER Working Paper 20511. Retrieved from http://www.nber.org/papers/w20511. Cambridge, MA: National Bureau of Economic Research.

Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The Effects of Education Accountability on Teachers: Are Policies Too-Stress Provoking for Their Own Good? *International Journal of Education Policy and Leadership, 4*(5), 1-13.

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping Stones: Principal career paths and school outcomes. *Social Science Research, 41*, 904-919.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals*. NBER Working paper No. 17803. Retrieved from http://www.nber.org/papers/w17803. Cambridge, MA: National Bureau of Economic Research.

Brewer, D. J. (1993). Principals and student outcomes: Evidence from U.S. high schools. *Economics of Education Review*, *12*(4), 281-292.

Brown, P. F. (2006). Preparing principals for today's demands. *Phi Delta Kappan*, *87*(7), 525-526.

Center on Education Policy. (2006). *Ten big effects of the No Child Left Behind Act on public schools*. Retrieved from http://www.cep-dc.org. Washington, DC: Center on Education Policy.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418-446.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252-279.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, *37*(4), 696-727.

Educational Testing Service (2008). *Standards, accountability, and flexibility: Americans speak on No Child Left Behind.* Princeton, NJ: Author.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3-28.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433-444.

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers and their work environment. *Educational Evaluation and Policy Analysis, 36*(4), 417-436.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*(2), 157-191.

Hargreaves, A., & Fink, D. (2006). *Sustainable leadership*. San Francisco, CA: Jossey Bass.

Koyama, J. (2014). Principals as bricoleurs: Making sense and making do in an era of accountability. *Educational Administration Quarterly, 50*(2), 279-304.

Lauren, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis, 34*(2), 185-208.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning: A review of research for the Learning from Leadership Project*. New York, NY: Wallace Foundation.

Li, D. (2012). *School accountability and principal mobility: How No Child Left Behind affects the allocation of school leaders*. Working paper. Harvard University.

Lyons, J. E., & Algozzine, B. (2006). Perceptions of the impact of accountability on the role of principals. *Education Policy Analysis Archives, 14*(16), 1-16.

Malani, A., & Reif, J. (2010). *Accounting for anticipation effects: An application to medical malpractice tort reform.* NBER Working Paper No. 16593. Cambridge, MA: National Bureau of Economic Research.

Mertler, C. (2011). Teachers' Perceptions of the Influence of *No Child Left Behind* On Classroom Practices. *Current Issues in Education, 13*(3). Retrieved on July 20, 2014, from http://cie.asu.edu/

Meyer, M. J., & Macmillan, R. B., Northfield, S. (2009). Principal Succession and Its Impact on Teacher Morale. *International Journal of Leadership in Education, 12*(2), 171-185.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review, 36*(1), 60-72.

National Research Council. (2011). *Incentives and Test-Based Accountability in Public Education*. Committee on Incentives and Test-Based Accountability in Public Education, Michael Hout and Stuart W. Elliott, *Editors*. Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management, 26*(4), 909-949.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy, 6*(3), 207-241.

Reback, R., Rockoff, J., Schwartz, H. L., & Davidson, E. (2011). Barnard/Columbia No Child Left Behind Database, 2002-03 and 2003-04. http://www.gsb.columbia.edu/nclb.

Salazar, K. (2007). *Senator Salazar's No Child Left Behind survey results*. Retrieved on September 17, 2015, from http://salazar.senate.gov/other/nclbresults.html#who.

Santoro, D. A. (2011). Good teaching in difficult times: Demoralization in the pursuit of good work. *American Journal of Education, 118*(1), 1-23.

Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom realities and No Child Left Behind*. The Civil Rights Project. Cambridge, MA: Harvard University.

United States Department of Education. (2001). *NCLB: Executive summary*. Retrieved on May 30, 2015, from http://www2.ed.gov/nclb/overview/intro/execsumm.html.

Wallace Foundation. (2013). *The school principal as leader: Guiding schools to better teaching and learning*. Washington, DC: Author.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced Leadership: What 30 years of research tells us about the effect of leadership on student achievement* (Working paper). Aurora, CO: Mid-continent Research for Education and Learning.

CHAPTER 2

INFLUENCE OF NCLB: A REVIEW OF LITERATURE, THEORIES BEHIND THE NCLB
ACCOUNTABILITY SYSTEM, AND AN EXAMINATION OF THE SYSTEM

Introduction

School accountability has been the center of debate among politicians, education

policymakers, and researchers in the nation since the 1990s, when a number of state and local

education agencies introduced school accountability systems in their K-12 education systems

(Figlio & Loeb, 2011; Linn, 2000). It is a performance-based incentive policy, under which

policymakers expect educators to change their instructional methods and daily practices to

achieve educational goals, such as raising student test scores and closing achievement gaps

among student subgroups. For example, Texas Education Agency (TEA) implemented an

accountability rating system in 1994, which is one of the strongest state accountability systems in

the nation. The TEA accountability system rates schools and districts based on student

performance on the State of Texas Assessments of Academic Readiness and then determines

their accreditation status (Texas Education Agency, 2014). Low-performing schools and districts

face interventions by local education agencies or TEA.

Among all kinds of school accountability systems, the most famous and influential

accountability system is the federal No Child Left Behind (NCLB) Act of 2001. The major goals

of NCLB were to raise the performance of all students in public schools and to ensure that no

students were trapped in persistently low-performing schools (U.S. Department of Education,

2001). To achieve these goals, policymakers created a sanction system under NCLB and

provided incentives for educators and district administrators. Some of the sanctions included

staff replacement, school closure, and state takeover. The basic assumption behind this system was that educators and administrators would respond to this sanction-based incentive and change their daily practices to meet the annual achievement goals.

A number of studies have investigated this assumption and analyzed the impact of NCLB both qualitatively and quantitatively. Some studies have found positive impacts on student achievement and other academic outcomes (e.g., Dee & Jacob, 2011; Lauren & Gaddis, 2012; Reback, Rockoff, & Schwartz, 2014); other studies argue that NCLB made teachers' daily work more stressful and challenging and caused unintended consequences (Berryhill, J., Linney, J. A., & Fromewick, J., 2009; Sunderman, Tracey, Kim, & Orfield, 2004; Valli & Buese, 2007). Most of these studies, however, have focused on student outcomes and teachers and, surprisingly, fewer studies have examined the impact on school principals. Little is known about how school principals responded to NCLB.

Lack of knowledge about principals' responses may be a problem in designing a new accountability system because, as the school leadership literature suggests, school principals play an important role in improving student and school outcomes (e.g., Branch, Hanushek, & Rivkin, 2012; Hallinger & Heck, 1998; Waters, Marzano, & McNulty, 2003). They serve as instructional leaders for teachers, and help them grow professionally and raise student achievement (e.g., Grissom & Loeb, 2011; Grissom, Loeb, & Master, 2013; Marks & Printy, 2003; O'Donnell & White, 2005; Sebastian & Allensworth, 2012). They are also important school managers, responsible for hiring and retaining teachers, building partnerships with parents and local community organizations, communicating with district administrators, and improving school climate, among others (Baker & Cooper, 2005; Brewer, 1993; Grissom & Loeb, 2011; Sanders & Harvey, 2002; Sebastian & Allensworth, 2012). This suggests that a school accountability

system that is not adequately designed to motivate school principals may backfire and cause unintended consequences. For example, principals may shift school resources from low-stakes subjects (e.g., music, art, and physical education) to high-stakes subjects (e.g., math and reading) and advise teachers to focus on students near the proficiency cut score. They may collaborate with teachers to cheat on high-stakes state assessments. Or, they may leave their schools to avoid sanctions.

This chapter provides a review of prior studies on the effects of NCLB on students, teachers, and principals, offers an analytical framework to examine the design of the NCLB accountability system, and assesses whether the design was appropriate. It proceeds as follows: The next section briefly describes the NCLB sanction system. It then reviews both qualitative and quantitative studies as well as opinion essays and survey reports; summarizing what we know about the effects of NCLB on students, teachers, and principals. After that, I discuss theories of incentives and motivations and provide an analytical framework to assess the design of the NCLB sanction system. The final section discusses implications for school principals' labor markets.

Mechanism of NCLB's AYP and Sanction Determinations

Under the NCLB system (NCLB, 2002), all public schools were required to make an adequate yearly progress (AYP) in reading and math in grades 3 to 8, and once at high school for *all* student subgroups.[1] They were also required to meet academic goals in one academic indicator (e.g., science assessment, attendance rates for elementary and middle schools and/or

---

[1] It typically included the following subgroups: all students, non-Hispanic White, non-Hispanic black, Hispanic, Asian/Pacific Islander, American Indian or Alaskan Native, multiracial, free/reduced lunch eligible, English language learners, and Individualized Education Program. A small number of states also used additional or different subgroups. For example, California used Filipino as a racial subgroup.

graduate rates for high schools).[2] Moreover, schools had to meet a 95 percent participation rate requirement for all student subgroups. To be eligible for AYP determination criteria, each subject-by-subgroup cell had to meet a cell-size requirement, which varied from state to state.

AYP for the two academic subjects was measured by the percent of students who performed at or above the proficiency level in state assessment(s). Each state education agency was given the discretion to define the proficiency level and set the annual measurable objectives (AMOs), which had to increase to 100 percent by the end of the 2013-14 school year.[3] Only Title I schools were subject to NCLB sanctions.

NCLB sanctions were progressive and cumulative. If a Title I school missed AYP for two consecutive years in the same subject and/or in the same additional indicator, NCLB required its district to offer public school choice (PSC) with transportation to all eligible students, so that they can transfer, if they prefer, to different public or charter schools that had not faced any sanction yet (USDOE, 2002). The district also had to provide the school with technical assistance to help them improve their student performance. If the school continued to miss AYP, the district had to offer supplemental educational services to students from low-income families, in addition to PSC and technical assistance. The severity of the sanction increased notably after missing AYP for four consecutive years. From this stage, the district had to take one of the corrective actions, some of which included staff replacement and internal reorganization of the school. If the school still continued to fail to meet AYP, it faced the planning and implementation of school restructuring. For any sanction to be lifted, it needed to make AYP for two consecutive years.

---

[2] The US Department of Education (USDOE) had required the additional academic indicator since the 2005-06 school year. Some states used writing or science assessment results as an additional academic indicator.

[3] The discretion caused wide variation in the definition of the proficiency level across the country. Students who performed at the proficient level in a state did not necessarily pass the proficiency level in another state. To understand the achievement levels of students across the country, Banderia de Mello (2011) compared the proficiency standards of the National Assessment of Educational Progress (NAEP) with those of states' assessments. He found that there exists a wide variation in the proficiency standards across the states, and that most states' proficiency standards are at or below NAEP's basic performance standards in both reading and math.

Making AYP for one year only delayed the sanction progression, and the school still faced the same sanction. A summary of the sanction system is presented in Table 1.

The sanction system allowed three exceptions in AYP determinations. The first exception was called the Confidence Interval. Many of the state education agencies that used this option calculated the upper bound of either a 95 or 99 percent confidence interval for the percent of proficiency achieved in failed subject(s) (USDOE, n.d.-a).[4] If a Title I school missed the target but the upper bound of the performance was equal to or exceeded the target, it was allowed to make AYP. The Confidence Interval was used to increase the reliability of AYP determinations. The second exception was called the Safe Harbor. It allowed schools which missed the annual target to make AYP if they successfully decreased the percent of students performing below the proficiency level by 10 percent from the previous year (USDOE, n.d.-b).[5] Some states also applied the Confidence Interval to the Safe Harbor. The confidence level ranged from 75 percent to 99 percent. The third exception was the Growth Model. Not all states were approved by the USDOE to use growth models for making AYP determinations (USDOE, n.d.-c). It incorporated into AYP determinations the number of students who performed below the proficient level but were "on track" to meeting their individual growth targets. More specifically, it added the number of students on track to the number of students performing at or above the proficient level.

---

[4] Some states applied the confidence interval to the annual target, not the actual percent of proficiency achieved. The confidence interval was usually applied to both the all-student group and student subgroups but a couple of states applied it to subgroups only. In addition, a few states used the confidence interval to small schools or small subject-by-subgroup cells only.

[5] The Safe Harbor was typically applied to student subgroups. Yet, a number of states used it to the all-student group as well.

Prior Research on the Influence of NCLB

Fourteen years since the implementation of NCLB, a large number of qualitative and quantitative studies have investigated the impact of NCLB. While a majority of studies have focused on the impact on student achievement, other studies explored how NCLB and its sanction system impacted school personnel including teachers and school principals. This section provides a general review of the literature on the impact of NCLB on students, teachers, and principals.

Influence of NCLB on Student Outcomes

Prior studies generally suggest that NCLB raised student achievement in both math and reading with more notable gains in math. For example, Lauen and Gaddis (2012) analyzed student achievement data in grades 3 through 8 in North Carolina from 2000 to 2008 to estimate the effect of subgroup-specific accountability threats. They find positive effects on minority and disadvantaged students, with the lowest performing schools posting the largest gains. Another study that used state assessment data across the nation finds that students in schools facing strong sanction threats performed at least as good as, or better than students in comparable schools that did not face such threats (Reback et al., 2014). Similarly, Springer (2008) finds that NCLB's sanction threat was positively correlated with test score gains among low-performing students. More recent studies that employed a regression discontinuity design using state assessment data also find that NCLB raised student achievement with variations in its impacts across the subjects and grades (Ahn & Vigdor, 2014; Chakrabarti, 2014; Fruehwirth & Traczynski, 2013; Hemelt, 2011).

These positive findings based on state assessment, generally concur with findings based on the National Assessment of Educational Progress (NAEP). Two studies have examined NAEP data before and after the implementation of NCLB by utilizing variation in the pre-NCLB state accountability systems across the states and found that NCLB improved student test scores in math (Dee & Jacob, 2011; Wong, Cook, & Steiner, 2009). A more recent study, which modified a statistical model used in the previous two studies, finds that NCLB improved reading test scores but not math test scores (Lee & Reeves, 2012). The consistency of these results provides a strong support for a general conclusion that NCLB positively impacted student achievement.

These positive gains may or may not result from a strategy that teachers reallocated more time and efforts to high-stakes subjects and to students near the AYP proficiency cut scores set by state education agencies. Researchers investigated the differential effects of NCLB based on the student position in the distribution of prior test scores. Their findings are inconsistent. Some researchers find that test score gains were achieved at the expense of top-performing and extremely low-performing students, whereas others suggest that there is no evidence on such "educational triage" (e.g., Ahn & Vigdor, 2014; Ballou & Springer, 2011; Booher-Jennings, 2005; Krieg, 2008; Neal & Schanzenback, 2007; Springer, 2008).

Influence of NCLB on Teachers

The previous section generally indicates that NCLB improved student performance. Because student learning mainly occurs in classrooms, it is unlikely that NCLB achieved its goals of raising student performance and narrowing achievement gaps without affecting the attitudes and instructional practices of classroom teachers (Grissom, Nicholson-Crotty, & Harrington, 2014). It is, therefore, reasonable to think that NCLB positively influenced teachers'

23

classroom practices. In fact, it is found that, under NCLB, teachers had higher clear academic expectations for student learning, which, in turn, increased their effort and commitment levels to help students learn, especially low-performing students, and thus worked longer hours (Finnigan & Gross, 2007; Guggino & Brint, 2010; Mintrop, 2004; Murnane & Papay, 2010; Reback et al., 2014). They also became more effective in planning and organizing lessons so that their lessons were aligned with the standards set by NCLB, engaged students more, and met the needs of low-performing students (Byrd-Blake, Afolayan, Hunt, Fabunimi, Pryor, & Leander, 2010; Dee, Jacob, & Schwartz, 2013; Guggino & Brint, 2010; Hamilton, Stecher, Marsh, McCombs, Robyn, Russell, Naftel, & Barney, 2007; Mintrop, 2004). Their level of classroom control, therefore, also increased (Grissom et al., 2014).

Although NCLB appears to have changed teachers' instructional practices positively, many teachers actually perceived NCLB negatively (Cawelti, 2006; Center on Education Policy, 2006; Deniston & Gerrity, 2010; Educational Testing Service, 2008; Guggino & Brint, 2010; Mintrop, 2004). One survey study reports that more than three-quarters of teachers (77 percent) perceived NCLB negatively (Educational Testing Service, 2008). Another study shows that a higher percentage of teachers (84 percent) reported unfavorable views toward NCLB (Guggino & Brint, 2010). Teachers claimed that, because of the annual performance targets set in math and reading, their role expectations increased (Valli & Buese, 2007). They had to shift their instructional time and resources from low-stakes subjects to high-stakes subjects and focused on "teaching to the test" in these subjects (e.g., Cawelti, 2006; Hamilton et al., 2007; Hannaway & Hamilton, 2008; Sunderman et al., 2004). In some cases, teachers and principals chose to exclude English language learners and students with disabilities from the state assessment requirements (Booher-Jennings, 2005; Cawelti, 2006).

Although these adjustments in instructional practices appears to have resulted in improvements in student achievement, NCLB gave teachers an increasing amount of pressure, which raised their stress level, and in some cases, led them to burn out (e.g., Berryhill et al., 2009; Center on Education Policy, 2006; Dee et al., 2013; Mertler, 2011; Santoro, 2011; Sunderman et al., 2004). This is especially the case if teachers internalized NCLB and found that it conflicted with their professional philosophy or standards (Mintrop, 2004). Such poor working conditions caused demoralization and reduced the commitment level and enthusiasm among teachers (e.g., Byrd-Blake et al., 2010; Cawelti, 2006; Center on Education Policy, 2006; Guggino & Brint, 2010; Mintrop & Trujillo, 2005; Santoro, 2011). Teachers working under such working conditions felt less secure about their job, and eventually decided to leave their schools (Deniston & Gerrity, 2010; Feng, Figlio, & Sass, 2010; Reback et al., 2014; Santoro, 2011; Sunderman et al., 2004).

These negative perspectives among teachers may suggest that NCLB needed to provide more support for them to continue teaching and commit to excellence and equity in education. In fact, teachers claimed that they needed more resources (e.g., time and money) to meet NCLB standards (Sunderman et al., 2004). They needed additional time to collaborate with other teachers to raise student performance.

Influence of NCLB on Principals

As briefly mentioned earlier, despite the crucial role that school principals play in improving school and student performance, fewer studies have systematically examined the impacts of NCLB on school principals. One survey study reports that 63 percent of principals had an unfavorable view toward NCLB (Educational Testing Service, 2008). Principals, in

general, perceived NCLB's testing requirements and sanction system negatively and claimed that the system was unfair (Brown, 2006; Lyons & Algozzine, 2006). One principal argues that NCLB added more complex responsibilities to the already challenging job (Brown, 2006). For example, principals had to spend more time on monitoring data on student achievement and preparing for NCLB's data reporting requirements (Koyama, 2014). Moreover, they did not receive adequate support from the state or local education agencies to comply with the NCLB requirements (Brown, 2006). These studies and reports generally suggest that NCLB impacted principals negatively. Yet, little research has empirically examined to what extent NCLB affected principals' leadership behaviors, working conditions, or daily practices.

A few recent studies have empirically investigated the impact of NCLB on principal turnover. Li (2012) has examined whether NCLB affected the distribution of principal quality using longitudinal administrative data from North Carolina. Using principal value-added scores, she finds that NCLB lowered the average principal quality at schools enrolling students from disadvantaged backgrounds. Principals with higher value-added scores are more likely to transfer to schools with lower probabilities of facing NCLB sanctions. In other words, NCLB shifted the distribution of principal quality across schools by affecting principal mobility and hiring behaviors. Ahn and Vigdor (2014) has examined the relationship between principal turnover and school restructuring under the NCLB sanction system using similar longitudinal administrative data from North Carolina, though the focus of the study was on student outcomes. As fewer schools faced school restructuring, their study examined whether principals exited the public education system after facing the sanction, instead of whether they transferred to different schools. They find that the sanction increased the probability of principal exit by 6 to 18 percentage points, although the impact was estimated imprecisely.

The previous section reviewed the *overall* impacts of NCLB. One of the key features of NCLB was its progressive sanction system. Because the severity of each sanction differs, it is possible that educators responded to each sanction differently. This section reviews studies about the impacts of each NCLB sanction.

District Technical Assistance

It is not widely known that the NCLB sanction system required school districts to provide technical assistance for low-performing schools facing any stage of sanction. Technical assistance included such supports as a provision of professional development for principals and teachers, and a suggestion for instructional strategies and methods of instruction (US Department of Education, 2002). Districts had to provide any support that had been proved effective based on credible scientific research.

Although district technical assistance was included as part of the NCLB sanction system, it may not have been considered as a sanction in practice. Rather, as the name suggests, it was an additional support that districts offered to low-performing schools. As a result, teachers and principals may have not necessarily felt a threat toward it. Unfortunately, little research has explored how educators perceived districts' technical assistance under the NCLB system or to what extent it helped them improve student performance. Although not directly relevant to the NCLB sanction system, one study about low-performing schools on probation under state accountability systems finds that some principals tended to have positive views toward the probation status because districts provided them with additional resources to transform schools (Mintrop, 2004). The principals claimed that they would not have had access to these resources

and assistance without the probation status. Although it is possible that principals viewed this district assistance as districts' invasion into their schools, the number of such principals might be small.

School Improve Year 1: School Choice

The first year of the NCLB sanction required school districts to provide technical assistance for Title I schools failing to make AYP for two consecutive years. Districts were also required to offer school choice to eligible students enrolled in failing schools. A handful of studies have examined the impacts of the NCLB-related school choice on student achievements. Recent studies, using a regression discontinuity design, find that the school choice threats raised student achievement (Ahn & Vigdor, 2014; Chakrabarti, 2014). Springer (2008) also finds positive impacts of the choice threat. Conversely, other recent studies using a similar regression discontinuity design suggest that schools that missed AYP for the first time did not make higher gains, or performed worse in the following year (Fruehwirth & Traczynski, 2013; Hemelt, 2011). Another study also finds no effect (Zimmer, Gill, Razquin, Booker, & Lockwood, 2007).

Few studies have examined principals' responses to the school choice threat. Available evidence, not directly related to NCLB, provides inconsistent patterns. For example, Cannata (2011) finds that principals perceived little competition threat from charter schools unless they were located in close proximity. Even if they were, principals did not necessarily change their leadership behaviors or time allocation. On the other hand, Kasman and Loeb (2013) find quite the opposite pattern: principals competed for students and changed their behaviors due to competition. They responded to competition through outreach and advertisement so that the

information that families received became more appealing. The principals also changed the instructional programs and services for students.

School Improvement Year 2: Supplemental Educational Services

A limited number of studies have examined the impact of the supplemental educational services (SES) on student achievement. SES were free tutoring programs offered to students who were eligible for the federal free or reduced lunch program and enrolled in Title I schools that were placed in the second year of the NCLB sanction (Springer, Pepper, Ghosh-Dastidar, 2014). Providers included for-profit organizations, non-profit organizations, local community organizations, school districts, and colleges and universities. As was the case with school choice, prior studies on this sanction show inconsistent findings (e.g., Chatterji, Kwon, & Sng, 2006; Chicago Public Schools, Office of Research, Evaluation and Accountability, Office of Extended Learning Opportunities, 2007; Heinrich, Meyer, & Whitten, 2010; Heistad, 2007; Springer et al., 2014; Zimmer et al., 2007). For example, Heinrich et al. (2010) find that attending SES did not significantly change student achievement, whereas Springer et al. (2014) report consistently significant and positive average effects.

One potential reason for the inconsistent findings about the impacts of both the school choice and SES under the NCLB sanction would be the low participation rates among eligible students. This causes a selection issue. One study finds that less than one percent of eligible students used a school choice option, whereas about a quarter of eligible students participated in supplemental educational services (Zimmer et al., 2007). This may be due to lack of information disseminated to parents, language barriers, and/or inadequate transportation provided (Burch, 2007). In addition, for SES, the quality of instructional services provided varied from provider to

provider (Heinrich et al., 2010), and the implementation of this service imposed additional administrative burdens on school districts (Burch, 2007). These factors likely contributed to the inconsistency in the findings among these studies.

As discussed for district technical assistance, SES could have been also viewed as a support for school improvement. It is likely that principals perceived this sanction positively. Yet, again, little research has examined how school principals responded to this sanction.

School Improvement Year 3 and Beyond

The severity of NCLB sanctions changed substantially from the third year. From this sanction stage, the sanction system required districts to take one of the corrective actions specified by the law, which included a replacement of staff responsible for low performance and internal reorganization of the school. This means that school principals might have started facing a risk of losing their job. This risk increases as the sanction stage elevates to the planning (Year 4) and implementation (Year 5 and beyond) of school restructuring. School restructuring plans included state takeover, contracting with a private education management company, closing and reopening as charter schools, reconstituting schools, and any other restructuring options such as hiring turnaround specialists, which was the most popular option that districts chose (Center on Education Policy, 2008a; Mathis, 2009).

Although there were multiple options that school districts could take to restructure low-performing schools, the available evidence generally suggests that school restructuring (including corrective actions taken by school districts) did not improve student achievement (e.g., Center on Education Policy, 2008b; Gill, Zimmer, Christman, & Blanc, 2007; Mathis, 2009). A more recent study using a regression discontinuity design finds that, although there appear to be

no significant effects of intermediate sanctions, school restructuring raised student achievement (Ahn & Vigdor, 2014).

Few studies have explored how principals perceived or responded to school restructuring under the NCLB sanction system. One report shows that principals often struggled with recruiting and hiring new effective teachers when many teachers were laid off (Center on Education Policy, 2008a). They started the new school year with many substitute teachers. This would have imposed more administrative burdens on principals. In addition, school restructuring may have increased principals' job pressure and stress levels because their job was not secured at this sanction stage. This highly stressful working environment may have led principals to depart voluntarily, as Ahn and Vigdor (2014) find. On the other hand, district administrators could have been dissatisfied with principals' job performance and replaced them with more effective leaders.

Theories of Motivations and Incentives behind the NCLB Sanction System

The NCLB sanction system was a performance-based incentive policy. If the incentive system is appropriately designed, policymakers can achieve their goals by motivating principals and other school staff to change their behaviors. If it is poorly designed, the policymakers would not be able to meet their goals, or may even face unintended consequences. This section discusses theories of motivations and incentives and offers an analytical framework to analyze the design of the NCLB sanction system and help understand how educators would respond to the sanction incentive.

In the following subsections, I first provide a general overview of the theories of motivations, incentives, and human behaviors. Thereafter, I discuss problems inherent in the

general incentive systems. After that, I discuss how to design an appropriate incentive system and use it to analyze the NCLB sanction system.

Motivations and Incentives

There are broadly two kinds of motivations that are important to distinguish to understand how incentive policies change human behaviors. One is intrinsic motivation, and the other is extrinsic motivation.[6] Intrinsic motivation is motivation that arises from within individuals. It is the desire a person has to meet his or her physiological and/or psychological needs through an engagement in an activity. Ryan and Deci (2000) define intrinsically motivated behaviors as "the doing of an activity for its inherent satisfactions rather than for some separable consequence" (p. 56). Eating and drinking are typical examples of intrinsically motivated behaviors. Learning mathematics would be also an intrinsically motivated behavior if a child is interested in learning it. Teaching would be another intrinsically motivated behavior if a teacher likes teaching and gains satisfaction and pleasure through the activity. Similarly, leading a school would be an intrinsically motivated behavior if a principal likes being a school leader and guiding teachers and other school administrators. Intrinsically motivated behaviors, therefore, should be "enjoyable, purposive, and provide sufficient reason to persist" (Pinder, 2011, as cited in Cerasoli, Nicklin, & Ford, 2014, p.981). Physiologically, a person engages in these intrinsically motivated behaviors when he or she is in a state of drive and external incentive is

---

[6] This may be an oversimplification of the concept of motivation. Ryan and Deci (2000) define motivation under the Self-Determination Theory. According to their taxonomy of human motivation, there are broadly three kinds of motivation: amotivation, extrinsic motivation, and intrinsic motivation. Amotivation is the state of no interest or no intention to act. Within extrinsic motivation, they classify motivation into four groups based on the degree of autonomy one possesses over his or her behavior. The groups include external regulation, introjection, identification, and integration. Ryan and Deci discuss that, the more autonomy or self-determination one has, the closer to intrinsic motivation his or her extrinsic motivation becomes.

present (Toates, 1986, 1994, as cited in Berridge, 2001). Because of these features, intrinsic

motivation would be innate or created through learning (White, 1959).

Intrinsic motivation is *enhanced* when a person feels competency and autonomy during

activities (Ryan & Deci, 2000). Cognitive Evaluation Theory argues that, when a person has

feelings of competence and autonomy during intrinsically motivated activities, he or she gains

more satisfaction of the need for competence and autonomy (Ryan & Deci, 2000). For example,

if a person is given *optimal* challenges with greater discretion over his or her approaches and

receives positive feedback from his or her supervisor, the person's intrinsic motivation will be

significantly enhanced. On the other hand, if a person faces challenges that are clearly

unachievable with his or her abilities and later receives negative feedback, his or her intrinsic

motivation will be diminished.

The impact of intrinsic motivation on performance hinges on how performance is defined

(Cerasoli et al., 2014). If it is defined based on performance *quality*, the predictive power of

intrinsic motivation is stronger because tasks that emphasize quality tend to be more complex

and require a higher level of engagement of technical skills and commitments. If intrinsically

motivated, a person is more likely to perform these tasks well. On the other hand, if performance

is defined based on performance *quantity*, the predictive power is weaker, because tasks

emphasizing quantity tend to be less complex and monotonous, which require a lower level of

engagement commitments.

Extrinsic motivation is often associated with incentives but can be defined as the desire to

obtain rewards or outcomes that are separable from activities that a person engages in (deCharms,

1968; Ryan & Dec, 2000). Incentives are "plans that have predetermined criteria and standards,

33

as well as understood policies for determining and allocation rewards" (Greene, 2011, p. 219).[7]

Rewards include money, grades, awards, recognition, praise, health benefits, and others (Cerasoli et al., 2014). For example, if a student engages in classroom activities, or studies hard to score high on a math exam in order to gain higher grades, he or she would be extrinsically motivated. If a teacher allocates more time to teach high-stakes subjects covered by a state or local assessment and less time to low-stakes subjects in order to gain financial rewards (e.g., bonuses) or avoid probation and sanctions, the teacher would be also extrinsically motivated.[8] As these examples suggest, extrinsically motivated behaviors are influenced by the expectancy of gains and losses (Cerasoli et al., 2014). It is important to note that incentives are provided under the assumption that people will change their behaviors and engage in activities so that they achieve outcome targets under the incentive plans and gain rewards or avoid sanctions (Greene, 2011).

Extrinsic motivation varies greatly in the degree of autonomy provided for individuals (Ryan & Deci, 2000). An incentive program may be prescriptive and does not offer much autonomy to participants. Or it may offer some autonomy and participants perceive incentivized outcomes or goals as personally important behaviors and accept the incentive plan as their own regulations. Whether an incentive plan is autonomous or not, for the plan to work, a person needs to internalize and integrate values and behavioral regulations under the plan. With successful internalization and integration of the plan and more personal commitment level, he or she will more persistently and positively engage in incentivized activities with better quality (Ryan & Deci, 2000).

---

[7] Criteria, standards, or policies for determining rewards are sometimes not clear or fuzzy (Kreps, 1997). For example, although a person is intrinsically motivated to work at his or her organization, the person fears discharge, which gives him or her an incentive to increase the current effort level. In this sense, some of the intrinsically motivated behaviors may be a combination of a person's response to incentives with fuzzy criteria and intrinsic motivation.

[8] In these examples, I do not necessarily mean that the student or the teacher is not intrinsically motivated at all, especially for the teacher; rather, what I mean is that they are *more* extrinsically motivated.

Policymakers use incentives to extrinsically motivate workers to adjust their behaviors such that policymakers can achieve their desired goals. Workers, on the other hand, gain reward for their performance, or sometimes, avoid punishments that result from bad performance.

Incentives and Human Behaviors

Incentives work under the assumption that people respond to them in order to gain rewards or avoid sanctions. Yet, this assumption does not always hold. People need to be in drive state where they feel gaining rewards or avoiding sanctions satisfy their needs; otherwise, they are less likely to respond to incentives. This is a key condition for incentives to work and "drive states modulate the value of incentive stimuli" (Berridge, 2001, p.237).[9] This means that if a person is in a situation where he or she lives under poverty and needs money, for example, the value of financial incentives would likely increase and the person would respond to them given that he or she accepts the incentive plans. On the other hand, if the person lives comfortably with a decent salary, the value of financial incentives would decrease, or at least be less than when he or she lives under poverty. Thus, incentives and drive states are both necessary to create a motivational state and lead a person to change his or her behaviors (Toates, 1986 & 1994, as cited in Berridge, 2001).

A person's behavior is reinforced and repeated in the future when the person changes his or her behavior given incentives, and receives rewards or avoids sanctions afterward. This reinforcement occurs under the condition that incentives are continuously offered. This claim originally dates back to Thorndike's Law of Effect (1898, as cited in Berridge, 2001). This law suggests that, if behaviors are followed by satisfaction (e.g., gaining rewards), they would be

---

[9] Physiologically, a phenomenon where hedonic value changes with drive states is called alliesthesia (Berridge, 2001).

more likely to be repeated in the future (Galef, 1998). Although Thorndike's law was based on

animal experiments, it is also likely to hold and make sense for humans. For example, if a child

receives an allowance every time the child cleans his or her room, the child is likely to continue

cleaning his or her room, as long as this financial incentive continues.

What happens to behaviors once incentives are removed is an empirical question that

researchers from psychology, economics, and neuroscience have investigated for the last four

decades. An answer to this question would influence policymakers' and employers' decisions on

whether or not they introduce extrinsic incentive programs to motivate people. For example, if

intrinsically motivated behaviors are less frequently observed than they are before the incentives

are offered, it would be costly to introduce such incentives.[10,11] This phenomenon is called

"undermining effect" or "crowding-out effect" in psychology and economics (see Camerer,

2010; Cerasoli et al., 2014; Frey & Jegen, 2001; Ledford, Gerhart, & Fang, 2013). On the other

hand, if behaviors still continue to be observed with more frequencies, incentives would be an

effective policy or program to use.

Research results on this topic appear to be divided with more supporting evidence from

laboratory experiments and opposing evidence from large-scale studies. For example, an

evaluation of a classic field experiment finds that the intrinsic interest demonstrated by preschool

children in a drawing activity decreased after engaging in the activity under incentives (Lepper,

Greene, & Nishbett, 1973). Deci (1971) finds that college students engage in activities less

frequently when monetary incentives are removed. Several meta-analysis studies in psychology

also find supporting evidence on the undermining effect (Cerasoli et al., 2014; Deci, Koestner, &

---

[10] Subhuman species (e.g., monkeys and dogs) tend to continue their behaviors even after the removal of incentives (Fester & Skinner, 1957; Morse, 1966; Uhl & Young, 1967, as all cited in Deci, 1971).

[11] Behaviors caused solely by extrinsic incentives will completely disappear once the incentives are removed. Yet, in daily life, it is rare to find behaviors solely motivated by extrinsic incentives.

Ryan, 1999; Deci, Koestner, & Ryan, 2001). A group of neuroscientists investigated how brain activities of college students change before and after incentives are given and then removed (Murayama, Matsumoto, Izuma, & Matsumoto, 2010). They find significant decreases in activities in the anterior striatum, midbrain, and lateral prefrontal cortex after incentives are removed, which suggests that intrinsic motivation is undermined by incentives.

There are various possible reasons for the undermining effect. First, intrinsic motivation may be a response to fuzzy extrinsic incentives such as fear of discharge (Camerer, 2010; Kreps, 1997). So the removal of incentives leads to a reduction in intrinsic (or fuzzy extrinsic) motivation. Second, people may believe that their motivation for tasks or activities is solely caused by extrinsic incentives when they are given, even if it is not actually true (Deci et al., 1999). Thus, once the incentives are removed, their motivation also disappears. Alternatively, incentives may be controlling; that is, they give individuals strong impetus toward specified outcomes, which likely reduces intrinsic motivation (Ryan, Mims, & Koestner, 1983). Such incentives often reduce a person's autonomy, which then weakens their intrinsic motivation (Cerasoli et al., 2014; Ryan & Deci, 2000). Third, the content of incentives may provide a signal for individuals – that incentivized tasks may be difficult or that the individuals are currently working below the acceptable level – which can undermine intrinsic motivation (Bénabou & Tirole, 2003; Camerer, 2010; Gneeny, Meier, & Rey-Biel, 2011). Another account from neuroscience, which neuroscientists believe is the most promising explanation, suggests that the brain (more specifically, the striatum) processes rewards through relative comparison of the intrinsic value of achieving tasks with that of gaining rewards, and rescales each value based on that of the other, fictive outcomes, and other yardsticks (Camerer, 2010; Murayama et al., 2010; Padoa-Schioppa, 2009; Seymour & McClure, 2008; Tobler, Fiorillo, & Schultz, 2005). This

means that, if the relative value of reward is significant, the intrinsic value of achieving tasks becomes smaller. As a result, once incentives are removed, the intrinsic value is underestimated, which leads to a decrease in motivation.

It is important to note that much of the supporting evidence comes from experiments conducted in laboratory settings, which are far different from settings in actual daily lives. In addition, the undermining effect may be mainly applicable to people with high initial levels of intrinsic motivation (Kreps, 1997). Thus, while these studies provide useful directions for future large-scale studies, their external validity may be relatively weak. In fact, large scale studies tend to show opposing evidence. For example, an evaluation of a merit scholarship program implemented in schools in Kenya finds no evidence that this scholarship weakened the intrinsic motivation among students in the program (Kremer, Miguel, & Thornton, 2009). When people were paid for attending a gym, their attendance increased even after such incentives were removed (Charness & Gneezy, 2009).[12] A more recent study of over 600 white-collar workers found no evidence that the pay-for-individual-performance plans undermine the workers' intrinsic interest, which is operationalized as intrinsic task satisfaction (Fang & Gerhart, 2012).

One reason for this opposing evidence may be habit formation (Charness & Gneezy, 2009). People who are given incentives or who participate in incentive programs typically do not have a habit of engaging in activities that are incentivized because they do not have intrinsic motivation in them, they do not have time to participate in them, they are physically unable to engage in them, and/or they care about how other people observe them (Charness & Gneezy, 2009). However, when they are given incentives, they may develop intrinsic interest in incentivized activities and start fostering a habit of regularly engaging in them. Once the regular

---

[12] Part of the study also provides moderate evidence that incentives backfire when incentives are given to people who have already been attending the gym.

habit is formed, they continue to participate in the activities after the incentives are removed, at least for a while. Another reason would be that incentives boost a person's competence in his or her performance once the person receives rewards (Cerasoli et al., 2014). This increased confidence may increase intrinsic motivation.

As this brief review of the literature suggests, the current literature does not provide a definitive answer to this question of the undermining effect. The conflict may be due to differences in the design of incentive programs, difficulties associated with their implementation, and/or differences in the target population. Given the typical costs associated with the implementation of an incentive policy, the answer to this question is critical and further research is needed to understand this phenomenon.

Problems Inherent in General Incentive Systems

The implementation of the incentives or the incentives themselves often encounters some problems which are critical to understand for the success of the incentive programs. One problem is the undermining effect or the crowding-out effect, as discussed in detail earlier from multiple perspectives. Another problem is the information that incentives convey from employers to workers (Gneezy et al., 2011). This means that, when large incentives are offered for particular tasks, employers signal that the tasks are difficult and workers may be less willing to engage in the tasks. Large incentives also signal that the employers do not trust their workers' abilities (Bénabou & Tirole, 2003) or are not satisfied with their current performance. This may negatively influence workers' responses. In addition, large incentives could change social norm and/or reduce image motivation (see Gneezy at al., 2011).

Inappropriate choice of rewards or sanctions becomes a problem. As discussed earlier, a person needs to be in a drive state when incentives are given in order to influence his or her behaviors (Berridge, 2001). If rewards or sanctions do not reduce drive, then incentives would not necessarily work. For example, monetary rewards may not work if workers are satisfied with their current salaries or if the amount of the rewards is small, relative to their salaries or their financial needs. Similarly, sanctions may not work if workers get used to them and are not desperate to get out of the sanctions. Alternatively, if a lot of workers receive sanctions, they feel less motivated to change their behaviors to avoid further sanctions.

Unclear or ambiguous reward criteria may lead to corruption in performance measures among workers (Kreps, 1997; National Research Council, 2011).[13] Workers may report their performance measures that may not reflect what employers desire to achieve (see Milgrom & Roberts, 1988, as cited in Kreps, 1997). Even if the criteria are clearly stated, if incentivized outcome measures do not capture all dimensions of the job (called multi-tasking problem), unintended consequences result (Dixit, 2002; National Research Council, 2011; Podgursky & Springer, 2007). For example, once a person learns the criteria, he or she may focus on incentive-related activities to gain rewards or avoid sanctions (Berridge, 2001) and not engage in activities not related to incentives but maybe important for the overall performance. Teaching to the test is one example. Teachers and principals facing accountability pressure allocate more time to high-stakes subjects (i.e., math and reading) and teach techniques to perform well on state assessments at the cost of other subjects (e.g., music, physical education, and social studies) and topics not covered in the state assessments (Hannaway & Hamilton, 2008; Sunderman et al., 2004). Cheating among educators is another example of unintended consequences (Jacob & Levitt,

---

[13] Kreps (1997) also argues that, when criteria are unclear, workers make opportunity responses not prescribed by incentive programs, and such responses will benefit employers.

40

2003; see Severson, 2011, for cheating in Atlanta Public Schools District). Furthermore, if incentive criteria do not agree with workers' own goals or performance standards, they would not respond to incentives. Kamenica (2012) suggests the following five factors that affect these unintended consequences: contextual inference, signaling, loss aversion, dynamic inconsistency, and choking.

Design of Incentives

The success of an incentive policy or program depends on its design. A poorly-designed incentive policy may result in unintended consequences and/or no desirable responses from workers. An ideal incentive policy or program would be one that addresses as many problems discussed above as possible, and does not negatively influence other important outcomes that are not included in the policy or program. Below, I discuss some of the key features that need to be considered in designing an incentive policy or program.

First, performance measures used need to measure the true value of workers' productivity or performance (National Research Council, 2011). Finding performance measures that reflect workers' true productivity would be relatively easy when the performance of a worker is uni-dimensional. For instance, the performance measure for factory workers could be the number of products that they produce or the number of product defects they produce in a given time. It would be more difficult to find performance measures when the performance of a worker has multiple dimensions (Dixit, 2002; Podgursky & Springer, 2007), which is applied to most of the white-collar jobs, including teaching and school leadership jobs. This requires employers to define the goals that they want performance measures to reflect (National Research Council, 2011). Once an appropriate set of performance measures is found, it is recommended to assign

weights based on the degree to which they are aligned with workers' true productivity (Baker, 2002).

The type of performance measures used should be carefully considered, as it affects a motivation-performance link (Cerasoli et al., 2014). If a performance measure focuses on quality, incentivizing the performance would increase intrinsic motivation and likely produce desired outcomes. If a performance measure emphasizes quantity such as the percent of students performing at or above the proficient level in state assessments, then incentivizing the performance is less likely to stimulate a worker's intrinsic motivation and may result in marginal or superficial performance improvement without affecting the overall performance.

Setting an appropriate level of performance targets influences workers' extrinsic motivation. For example, if a performance target is set low, workers with high productivity would feel it easy to achieve that target and thus would be less motivated. On the other hand, if the target is set high, workers with relatively low productivity would view it as an unattainable target and, therefore, become less motivated. If the same performance target is used for everyone, high- and low-productive workers would not be motivated to respond to incentives (Lazear, 2000; National Research Council, 2011). As research suggests, "the highest level of effort occurs when the task is moderately difficult" (National Research Council, 2011, p2-13). Differentiated performance targets based on workers' productivity level may be another effective way to achieve the overall goals.

The size of incentives should be significant enough to motivate workers to respond to incentives (Goldhaber, 2007).[14] For example, if a financial reward is not large enough relative to a worker's current salary or their current needs, it may not motivate the worker to achieve

---

[14] As discussed earlier, large-size incentives convey some negative information. Yet, the benefits from large-size incentives are likely to exceed the costs associated with negative consequences.

performance targets. Such reward may, instead, worsen a worker's job performance. It may be better not to offer any rewards at all (Gneezy & Rustichini, 2000; Gneezy et al., 2011). Moreover, if outcome measures are influenced by uncertainty and uncontrollable factors, rewards need to be large enough to compensate for them (National Research Council, 2011). Similarly, sanctions need to be great enough and implemented as designed to adequately motivate workers.

The unit of incentives can be the individual, the team (or group), or the organization. This choice influences a worker's behaviors. If it is set at the individual level, it would be easier to measure a worker's contribution; yet this may cause competition among workers, and hence, a lack of collaboration (National Research Council, 2011). On the other hand, if it is set at the team or group level, team collaboration would be enhanced, but it becomes difficult to measure an individual worker's contribution to the team performance. At the same time, it may cause some free-riders. Choice of the unit depends on the relative importance of team cooperation and the degree of uncertainty associated with use of team-based performance measures (National Research Council, 2011).

Criteria used to determine rewards need to be clear to all participants so that they understand what levels of efforts are needed to gain rewards or avoid sanctions; otherwise, extrinsic motivation among workers will be weaker (Cerasoli et al., 2014). This is more about the implementation of the incentives, but if incentive programs are not clearly explained to workers, they would not respond to incentives, not because incentives are not effective but because the information dissemination strategy is not effective.

Finally, the duration of incentive policies and programs needs to be long enough for an employer to be able to observe changes in workers' behaviors and improvements in performance outcomes. Short incentive programs may be completed before people start adjusting their

43

behaviors or forming habits. Yet, as discussed earlier, it is important to remember that the removal of incentive programs, whether they are short or long, may reduce intrinsic motivation among workers.

Analysis of the Design of the NCLB Sanction System

As described earlier, schools under the NCLB sanction system were required to meet AYP set by each state education agency in math and reading for all students and subgroups. They also had to meet the test participation rate requirement and the additional academic indicator requirement. Missing AYP in the same subject or the same indicator for two consecutive years placed schools in the NCLB sanctions. Schools needed to make AYP for two consecutive years to get out of the sanctions.

The analytical framework from the theories of incentives and motivations suggests some problems in the NCLB sanction system, which may have hindered educators from changing their behaviors to meet policymakers' goals. First, performance measures in the system do not appear to have reflected true values of job performance among educators. A typical annual performance target was set as the percent of proficiency that schools needed to achieve. For example, the annual proficiency target in the state of Missouri was 42.9 percent in communication arts and 35.8 in math in the 2006-07 school year. It was 85.4 percent and 76.5 percent in Massachusetts, respectively. However, the performance of educators has multiple dimensions, and student achievement is one of the dimensions (Podgursky & Springer, 2007). For example, principals typically serve two roles: instructional leaders and school managers. Student achievement measures in math and reading could be used as the main job performance measures as an instructional leader; yet, they would not capture the other important job tasks not related to these

44

academic subjects such as creating programs to promote healthy diet, deepening an understanding of music and arts history, and encouraging regular physical exercises, let alone job performance as school managers. Managerial or administrative tasks include managing class schedules, fostering supportive relationships with parents and local community organizations, communicating with district and state administrators, hiring and retaining teachers, maintaining building facilities, and keeping safe school environments (Baker & Cooper, 2005; Brewer, 1993; Grissom & Loeb, 2011; Sanders & Harvey, 2002; Sebastian & Allensworth, 2012). All of these job tasks are equally important to operate schools. The NCLB sanction system did not consider these dimensions of principals' job tasks.

Along with this multi-tasking problem, these performance measures and sanction criteria may have conflicted with educators' professional standards and personal educational philosophies. For example, if a principal or a teacher believes that the purpose of education is not just to raise test scores, he or she may not internalize or integrate values that the NCLB sanction system promoted (Ryan & Deci, 2000). This would be more likely to be the case if they have greater public service motivation, which cultivates a sense of personal sacrifice and altruism over time to serve struggling students (Gregg, Grout, Ratcliffe, Smith, & Windmeijer, 2011; Levacic, 2009; Perry & Wise, 1990; Perry, 2000). In fact, studies report that a majority of educators viewed the system unfavorably (Educational Testing Service, 2008; Guggino & Brint, 2010).

Second, performance measures in the system emphasized quantity, not quality. This emphasis might give educators less autonomy over how to educate their students, which could weaken intrinsic motivation (Ryan & Deci, 2000). If educators are intrinsically more motivated to teach students, the emphasis on quantity would be less likely to motivate educators to change their behaviors to achieve the proficiency goals set by state education agencies (Cerasoli et al.,

45

2014) because this emphasis would conflict with their professional standards. Extrinsically more motivated educators, on the other hand, would focus on raising the number of students performing at or above the proficient level in high-stakes state assessments, and thus, shift their resources to high-stakes subjects and students "on the bubble" at the cost of other important educational goals. As mentioned earlier, teaching to the test and cheating are some examples of behaviors that result from the emphasis on quantity (Jacob & Levitt, 2003; Severson, 2011). Student performance may improve in the short run but it is marginal, temporal, and/or superficial, and would not lead to improvements in their overall performance. An emphasis on quality would be, therefore, more appropriate in education.

Third, state education agencies applied the same annual proficiency targets to all schools, although the Safe Harbor takes into consideration the growth trajectories in school performance over time. If school characteristics and performance are equally distributed within states, equal performance targets would work well. Yet, the existing literature suggests that this is not the case. Quality and qualifications of educators are unequally distributed across schools and districts within states (Clotfelter, Ladd, & Vigdor, 2005; Glazerman & Max, 2011; Goldhaber, Lavery, & Theobald, 2014; Grissom & Mitani, 2016; Isenberg, Max, Gleason, Potamites, Santillano, & Hock, 2013; Lankford, Loeb, & Wyckoff, 2002; Loeb, Kalogridges, & Horng 2010; Sass, Hannaway, Xu, Figlio, & Feng, 2010). As a result, for high-performing schools, annual proficiency targets were too easy to achieve; on the other hand, these targets were too difficult to meet for low-performing schools. Educators serving these low-performing schools might have felt unfair about the system (National Research Council, 2011). Educators serving these high- and low-performing schools may not have been motivated to respond to the NCLB sanction system.

Finally, under the NCLB sanction system, educators may not have perceived the first two years of the sanctions as sanctions. As described earlier, the first two years of the sanctions required districts to provide technical assistance for schools that missed AYP and offer school choice options and supplemental educational services among eligible students. Districts' technical assistance and supplemental education services were actually additional educational resources provided for failing schools for free. These two years may not have provided enough incentives for failing schools to improve their performance. Rather, these years may have motivated schools to fail until this sanction stage.

This analysis of the design of the NCLB sanction system suggests a possibility that the sanction system may not have worked as designed and caused unintended consequences. Prior studies about the impact of NCLB provide some evidence on this possibility.

Labor Market Implications

Problems in the design of the NCLB sanction system have some implications for school principals' labor markets. First, these problems are likely to have influenced school principals' working conditions. For example, principals may need to allocate more time on preparation for high-stakes state assessments such as math and reading and spend less time on other low-stakes academic and non-academic subjects. They may also have to work longer hours per week to ensure that low-performing students exceed a proficiency cut score. This pressure could have been stronger among principals working at schools around the AYP cut score. For these principals, accountability comes first before students (Knapp & Feldman, 2012). If principals can internalize and integrate values that the sanction system emphasized, these working conditions may not be a problem. Yet, if the system's performance criteria do not align with the principals'

professional standards or personal philosophies of education, these working conditions would raise the principals' job stress and dissatisfaction levels.

Second, high levels of job stress and dissatisfaction may have led to principal turnover, especially among principals serving schools near the AYP cut point. When job stress and dissatisfaction levels arise, school principals would be more likely to burn out and affect their well-being. They will start thinking about alternative jobs available to them including principalship positions and other school staff positions at different schools, district administrator positions, and positions outside the public education system. Principals conduct cost-benefit analyses by comparing their current working conditions and labor contracts with those of potential future jobs. If the net benefits of taking a new job exceed the net costs of staying in the current school, principals turn over. If not, they stay in their school. Because the accountability (or sanction) pressure would be greater among principals working at schools around the AYP cut score, these principals may be more likely to leave for schools far above the cut score or move into private schools. On the other hand, principals working at high-performing schools may be more likely to stay at their schools to protect them from federal interventions.

It is important to note that principal labor markets could be influenced by school district administrators as well because they were also subject to the NCLB sanction system. They would behave to meet district-level AYP targets. From their perspectives, one strategy would be to identify high-performing school principals and reassign them to failing schools, especially those schools near the AYP cut point. These principals are *pulled* by district administrators to these schools. On the other hand, district administrators may not renew a contract with principals if their schools persistently perform poorly on high-stakes state assessments. These principals are *pushed out*.

In sum, the current literature lacks empirical evidence on the influence of NCLB on school principals, and there is a need for research that investigates principal behaviors under NCLB using large-scale data. Given that NCLB was a performance-based incentive policy, its influence may need to be explored by using an analytical framework based on the theories of incentives and motivations.

References

Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina*. NBER Working Paper No. 20511. Retrieved from http://www.nber.org/papers/w20511. Cambridge, MA: National Bureau of Economic Research.

Baker, B. D., & Cooper, B. S. (2005). Do principals with stronger academic backgrounds hire better teachers? Policy implications for improving high-poverty schools. *Educational Administration Quarterly, 41*(3), 449-479.

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *The Journal of Human Resources, 37*(4), 728-751.

Ballou, D., & Springer, M. G. (2011). *Has NCLB encouraged educational triage? Accountability and the distribution of achievement gains*. Unpublished manuscript. Nashville, TN: Vanderbilt University.

Banderia de Mello, V. (2011). *Mapping state proficiency standards onto the NAEP scales: Variation and change in state standards for reading and mathematics, 2005-2009* (NCES 2011-458). National Center for Education Statistics, Institute of Education Statistics, U.S. Department of Education, Washington, DC: Government Printing Office.

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, *70*(3), 489-520.

Berridge, K. C. (2001). Reward learning: Reinforcement, incentives, and expectations. *The Psychology of Learning and Motivation, 40*, 223-278.

Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The Effects of Education Accountability on Teachers: Are Policies Too-Stress Provoking for Their Own Good? *International Journal of Education Policy and Leadership, 4*(5). Retrieved on July 19, 2014, from http://www.ijepl.org

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231-268.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals*. NBER Working paper No. 17803. Retrieved from http://www.nber.org/papers/w17803. Cambridge, MA: National Bureau of Economic Research.

Brewer, D. J. (1993). Principals and student outcomes: Evidence from U.S. high schools. *Economics of Education Review, 12*(4), 281-292.

Brown, P. F. (2006). Preparing principals for today's demands. *Phi Delta Kappan*, *87*(7), 525-526.

Burch, P. (2007). *Supplemental education services under NCLB: Emerging evidence and policy issues.* East Lansing, MI: The Great Lakes Center for Education Research & Practice.

Byrd-Blake, M., Afolayan, M.O., Hunt, J.W., Fabunmi, M., Pryor, B.W., & Leander, R. (2010). Morale of teachers in high poverty schools: A post-NCLB mixed methods analysis. *Education and Urban Society, 42*(4), 450-472.

Camerer, C. F. (2010). Removing financial incentives demotivates the brain. *Proceedings of the National Academy of Sciences, 107*(49), 20849-20850.

Cannata, M. (2011). *How do principals respond to charter school competition?* Nashville, TN: National Center on School Choice.

Cawelti, G. (2006). The side effects of NCLB. *Educational Leadership, 64*(3), 64-68.

Center on Education Policy. (2006). *Ten big effects of the No Child Left Behind Act on public schools*. Retrieved from www.cep-dc.org. Washington, DC: Center on Education Policy.

Center on Education Policy. (2008-a). *A call to restructure restructuring: Lessons from the No Child Left Behind Act in five states*. Washington, DC: Center on Education Policy.

Center on Education Policy. (2008-b). *Has student achievement increased since 2002? State test score trends through 2006–07*. Retrieved on November 14, 2014, from tttp://www.cep-dc.org

Cerasoli, C P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*(4), 980-1008.

Chakrabarti, R. (2014). Incentives and responses under No Child Left Behind: Credible threats and the role of competition. *Journal of Public Economics, 110*, 124-146.

Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica, 77*(3), 909-931.

Chatterji, M., Kwon, Y. A., & Sng, C. (2006). Gathering evidence on an after-school supplemental instruction program: Design challenges and early findings in light of NCLB. *Education Policy Analysis Archives, 14*(12), 1-44.

Chicago Public Schools: Office of Research, Evaluation, and Accountability, Office of Extended Learning Opportunities. (2007). *SES Tutoring Programs: An Evaluation of Year 3 in the Chicago Public Schools*. Chicago, IL: Chicago Public Schools.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J.L. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review, 24*(4), 377-392.

deCharms, R. (1968). *Personal causation: The internal effective determinants of behavior*. New York, NY: Academic Press.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105-115.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*, 627-668.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 71*(1), 1-27.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252-279.

Deniston, R. D., & Gerrity, K. W. (2010). Elementary school teachers' perceptions of No Child Left Behind and its effect on morale. *scholarlypartnershipedu*, *5*(2).

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, *37*(4), 696-727.

Educational Testing Service (2008). *Standards, accountability, and flexibility: Americans speak.* Princeton, NJ: Author.

Fang, M., & Gerhart, B. (2012). Does pay for performance diminish intrinsic interest? *The International Journal of Human Resource Management, 23*(6), 1176-1196.

Feng, L., Figlio, D., & Sass, T. (2013). *School accountability and teacher mobility*. Working paper. Atlanta, GA: Georgia State University.

Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbooks in Economics*, Vol. 3, 383-421.

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

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589-611.

Fruehwirth, J. C., & Traczynski, J. (2013). *Spare the rod? The dynamic effects of failing accountability on schools*. Unpublished manuscript. University of Cambridge.

Galef, B. G., Jr. (1998). Edward Thorndike: Revolutionary psychologist, ambiguous biologist. A*merican Psychologist, 53*(10), 1128-1134.

Gill, B., Zimmer, R., Christman, J., & Blanc, S. (2007). *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: RAND.

Glazerman, S., & Max, J. (2011). *Do low-income students have equal access to the highest performing teachers?* Retrieved on December 29, 2014, from http://ies.ed.gov/ncee/pubs/20114016/pdf/20114016.pdf. Washington, DC: National Center for Education Evaluation and Regional Assistance, U.S. Department of Education.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25*(4), 1-21.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, *115*(3), 791-810.

Goldhaber, D. (2007). *Principal compensation: More research needed on a promising reform*. Washington, DC: Center for American Progress.

Goldhaber, D., Lavery, L., & Theobald, R. (2014). Uneven playing field? Assessing the inequity of teacher characteristics and measured performance across students. Paper presented at the Association for Education Finance and Policy Annual Conference in San Antonio, TX.

Greene, R. J. (2011). *Rewarding performance: Guiding principles, custom strategies*. New York, NY: Routledge.

Gregg, P., Grout, P. A., Ratcliffe, A., Smith, S., & Windmeijer, F. (2011). How important is pro-social behavior in the delivery of public services? *Journal of Public Economics*, *95*, 758-766.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal, 48*(5), 1091-1123.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher, 42*(8), 433-444.

Grissom, J. A., & Mitani, H. (2016). *Principal sorting and the distribution of principal quality: Documenting inequities in state and national data*. Working paper. Nashville, TN: Vanderbilt University.

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers and their work environment. *Educational Evaluation and Policy Analysis, 36*(4), 417-436.

Guggino, P. C., & Brint, S. (2010). Does the No Child Behind Act help or hinder K-12 education? *Policy Matters*, *3*(3).

Hallinger, P. & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*(2), 157-191.

Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J. L., Naftel, S., & Barney, H. (2007). *Standards-based accountability Under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND.

Hannaway, J., & Hamilton, L. (2008). *Performance-based accountability policies: Implications for school and classroom practices*. Washington, DC: The Urban Institute.

Heinrich, C. J., Meyer, R. H., Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis*, *32*(2), 273-298.

Heistad, D. (2007). *Evaluation of supplemental education services in Minneapolis public schools: An application of matched sample statistical design*. Minneapolis, MN: Minneapolis Public Schools.

Hemelt, S. W. (2011). Performance effects of failure to make Adequate Yearly Progress (AYP): Evidence from a regression discontinuity framework. *Economics of Education Review*, *30*(4), 702-723.

Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., & Hock, H. (2013). *Access to effective teaching for disadvantaged students*. Washington, DC: National Center for Education Evaluation and Regional Assistance, U.S. Department of Education.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics, 118*(3), 843-877.

Kamenica, E. (2012). Behavioral economics and psychology of incentives. *Annual Review of Economics, 4*, 1-26.

Kasman, M., & Loeb, S. (2013). Principals' perceptions of competition for students in Milwaukee schools. *Education Finance and Policy, 8*(1), 43-73.

Knapp, M. S., & Feldman, S. B. (2012). Managing the intersection of internal and external accountability: Challenge for urban school leadership in the United States. *Journal of Educational Administration*, *50*(5), 666-694.

Koyama, J. (2014). Principals as bricoleurs: Making sense and making do in an era of accountability. *Educational Administration Quarterly, 50*(2), 279-304.

Kremer, M., Miguel, E., & Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics, 91*(3), 437-456.

Kreps, D. M. (1997). Intrinsic motivation and extrinsic incentives. *American Economic Review Papers and Proceedings, 87*(2), 359-364.

Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind Act. *Education Finance and Policy, 3*(2), 250–281.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37-62.

Lauren, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis, 34*(2), 185-208.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review, 90*(5), 1346-1361.

Ledford, G. E., Gerhart, B., & Fang, M. (2013). Negative effects of extrinsic rewards on intrinsic motivation: More smoke than fire. *WorldatWork*, *22*(2), 17-29.

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990-2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis, 34*(2), 209-231.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, *28*(1), 129-137.

Levacic, R. (2009). Teacher incentives and performance: An application of principal-agent theory. *Oxford Development Studies, 37*(1), 33-46.

Li, D. (2012). *School accountability and principal mobility: How No Child Left Behind affects the allocation of school leaders*. Working paper. Harvard University.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, *29*(2), 4-16.

Loeb, S., Kalogridges, D., & Horng, E. L. (2010). Principal preferences and the even distribution of principals across schools. *Educational Evaluation and Policy Analysis*, *32*(2), 205-229.

Lyons, J. E., & Algozzine, B. (2006). Perceptions of the impact of accountability on the role of principals. *Education Policy Analysis Archives, 14*(16), 1-16.

Marks, H. M., & Printy, S M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3), 370-397.

Mathis, W. J. (2009). *NCLB's ultimate restructuring alternatives: Do they improve the quality of education?* East Lansing, MI: The Great Lakes Center for Education Research & Practice.

Mertler, C. (2011). Teachers' Perceptions of the Influence of *No Child Left Behind* On Classroom Practices. *Current Issues in Education*, *13*(3). Retrieved on July 20, 2014, from http://cie.asu.edu/

Mintrop, H. (2004). *Schools on probation: How accountability works (and does not work)*. New York, NY: Teachers College Press.

Mintrop, H., & Trujillo, T. M. (2005). Corrective action in low performing schools: Lessons for NCLB implementation from first-generation accountability systems. *Education Policy Analysis Archives, 13*(48), 1-28.

Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences*, *107*(49), 20911-20916.

Murnane, R. J., & Papay, J. P. (2010). Teachers' views on No Child Left Behind: Support for the principles, concerns about the practices. *Journal of Economic Perspectives, 24*(3), 151-166.

National Research Council. (2011). *Incentives and Test-Based Accountability in Public Education*. Committee on Incentives and Test-Based Accountability in Public Education, Michael Hout and Stuart W. Elliott, *Editors*. Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics, 92*(2), 263–283.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat 1425 (2002).

O'Donnell, R. J., & White, G. P. (2005). Within the accountability era: Principals' leadership behaviors and student achievement. *NASSP, 89*(645), 56-71.

Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. *The Journal of Neuroscience, 29*(44), 14004-14014.

Perry, J. L. (2000). Bringing society in: Toward a theory of public-service motivation. *Journal of Public Administration Research and Theory, 10*(2), 471-488.

Perry, J. L., & Wise, L. R. (1990). The motivational bases of public service. *Public Administration Review*, *50*(3), 367-373.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management, 26*(4), 909-949.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy, 6*(3), 207-241.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54-67.

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology, 45*(4), 736-750.

Sanders, M. G., & Harvey, A. (2002). Beyond the school walls: A case study of principal leadership for school-community collaboration. *Teachers College Record*, *104*(7), 1345-1368.

Santoro, D. A. (2011). Good teaching in difficult times: Demoralization in the pursuit of good work. *American Journal of Education*, *118*(1), 1-23.

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2010). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics, 72*(2), 104-122.

Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on classroom instruction and student learning: A study of mediated pathways to learning. *Educational Administration Quarterly*, *48*(4), 626-663.

Severson, K. (2011, July 5). Systematic cheating is found in Atlanta's school system. *New York Times*. Retrieved on February 15, 2015, from http://www.nytimes.com/2011/07/06/education/06atlanta.html?_r=0.

Seymour, B., & McClure, S. M. (2008). Anchors, scales and the relative coding of value in the brain. *Current opinion in neurobiology*, *18*(1), 1-6.

Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, *27*(5), 556-563.

Springer, M. G., Pepper, M. J., & Ghosh-Dastidar, B. (2014). Supplemental educational services and student test score gains: Evidence from a large, urban school district. *Journal of Education Finance*, *39*(4), 370-403.

Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom realities and No Child Left Behind*. The Civil Rights Project. Cambridge, MA: Harvard University.

Texas Education Agency. (2014). *2014 accountability manual*. Retrieved on May 30, 2015, from http://www.easybib.com/reference/guide/apa/website.

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, *307*(5715), 1642-1645.

United States Department of Education. (2001). *NCLB: Executive summary*. Retrieved on May 30, 2015, from http://www2.ed.gov/nclb/overview/intro/execsumm.html.

United States Department of Education. (2002). *No Child Left Behind: A desktop reference 2002*. Washington, DC: United States Department of Education.

United States Department of Education. (n.d.-a). ED data express: Data about elementary & secondary schools in the U.S. Retrieved on November 10, 2014, from http://eddataexpress.ed.gov/data-element-explorer.cfm. Washington, DC: United States Department of Education.

United States Department of Education. (n.d.-b). ED data express: Data about elementary & secondary schools in the U.S. Retrieved on November 10, 2014, from http://eddataexpress.ed.gov/definitions.cfm. Washington, DC: United States Department of Education.

United States Department of Education. (n.d.-c). Growth models. Retrieved on November 10, 2014, from http://www2.ed.gov/admins/lead/account/growthmodel/index.html?exp=3. Washington, DC: United States Department of Education.

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal, 44*(3), 519-558.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced Leadership: What 30 years of research tells us about the effect of leadership on student achievement* (Working paper). Aurora, CO: Mid-continent Research for Education and Learning.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*(5), 297-333.

Wong, M., Cook, T. D., & Steiner, P. M. (2009). *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series*. Northwestern University Institute for Policy Research Working Paper. Evanston, IL: Northwestern University.

Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood, J. R. (2007). State and local implementation of the No Child Left Behind Act: Volume I – Title I school choice, supplemental educational services, and student achievement. Report to the U.S. Department of Education Office of Planning, Evaluation, and Policy Development.

Table 1: NCLB sanctions

Number of *consecutive* years that school missed AYP in same subject in at the end of year *t-1*	Sanction imposed in the *following* school year (year *t*)
0	No sanction imposed.
1	No sanction imposed but the failure is publicly announced.
2	District must offer Public School Choice (PSC) with transportation to all students to transfer to another public school or charter school within the same district which has not been identified for "in-need-of improvement. It also needs to provide technical assistance.
3	District must offer supplemental educational services (SES) to students from low-income families in addition to PSC and technical assistance.
4	District must take at least one of the following corrective actions on failing schools: (1) replace school staff, (2) implement a new curriculum (with appropriate professional development), (3) decrease management authority at the school level, (4) appoint an outside expert to advise the school, (5) extend the school day or year, and (6) reorganize the school internally. District continues to offer PSC, SES, and technical assistance.
5	District must initiate plans to fundamentally restructure failing schools. It continues to offer PSC, SES, and technical assistance.
6	District must implement the school restructuring plan. It includes one or more of the following actions: (1) reopen the school as a public or charter school, (2) replace all or most of the staff who are responsible for the failure to make AYP, (3) enter into a contract with an entity, such as private management company to operate the school, (4) turn the operation of the school over to the state department of education, and (5) other major restructuring. District continues to offer PSC, SES, and technical assistance.
7 and beyond	School remains in restructuring until it meets AYP for two consecutive years.

Notes: If a district does not have PSC options because all schools are identified for "in-need-of improvement", it offers SES instead. Data source: US Department of Education

CHAPTER 3

PRINCIPAL TURNOVER UNDER NCLB ACCOUNTABILITY PRESSURE

Introduction

The federal No Child Left Behind Act of 2001 (NCLB) was a performance-based accountability policy that was designed to motivate educators and administrators to change behaviors in order to improve school and student outcomes. Schools received sanctions if they did not meet adequate yearly progress (AYP) goals set by each state education agency.

A simplification of the logic behind this accountability policy was that educators and administrators would respond to sanctions. For example, principals and teachers may allocate more time to high-stakes subjects, such as math and reading. Teachers may change their instructional practices to help struggling students. They may also spend more time on interacting with students. Principals could assign effective teachers to a group of students from whom they expect immediate improvement in test scores.

A number of researchers have examined this assumption and, more broadly, the impact of NCLB. For example, studies find that NCLB, in general, raised overall student performance and the performance among minority and disadvantaged students (e.g., Ballou & Springer, 2011; Dee & Jacob, 2011; Lauen & Gaddis, 2012; Reback, Rockoff, & Schwartz, 2014). Other studies report that teachers viewed the NCLB system negatively, and claim that it increased the amount of pressure, which raised their job stress levels, and, in some cases, led them to burn out or become demoralized (e.g., Berryhill, Linney, & Fromewick, 2009; Center on Education Policy, 2006; Dee, Jacob, & Schwartz, 2013; Mertler, 2011; Santoro, 2011; Sunderman, Tracey, Kim, &

Orfield, 2004). A more recent empirical study finds that while NCLB negatively influenced teachers' perceptions of cooperation with their colleagues, it did not change their job satisfaction or commitment levels (Grissom, Nicholson-Crotty, & Harrington, 2014).

Although the primary focus of NCLB studies is on students and teachers, school principals were, in fact, a prime target of NCLB. They were required to make a good deal of important decisions to respond to NCLB sanctions, and, at a certain sanction stage, they were the target of replacement as a result of low performance. Anecdotal evidence suggests that NCLB made principalship more challenging and that many principals could not meet the growing demand of the law, which increased their job stress (Brown, 2006). Principals report that high-stakes accountability systems, in general, created a culture of fear among school leaders (McGhee & Nelson, 2005). In some cases, NCLB pressure drove the principals to tamper with test scores (Barry, 2015; Berman, 2015; Gabriel, 2010). Moreover, survey reports show that a majority of principals viewed NCLB unfavorably, and sometimes perceived it as an unfair system (Educational Testing Service, 2008; Lyons & Algozzine, 2006; Salazar, 2007). These stressful working conditions could have led to frequent turnover among principals, especially those working at low-performing schools, which had a higher probability of facing the NCLB sanctions (Ahn & Vigdor, 2014; Li, 2012; White & Agarwal, 2011).

Principal turnover, particularly repeated and frequent turnover within a school, has been found to cause conflict between the teachers and new principals, disruption of the instructional program, and changes in school vision, which are likely to delay and destabilize school reforms, and may negatively impact teacher morale (Hargreaves & Fink, 2006; Meyer, Macmillan, & Northfield, 2009). As a result, principal turnover could create a negative impact on student performance. Recent studies find that principal turnover, especially frequent principal turnover,

is associated with declines in student achievement and lower teacher retention (Béteille, Kalogrides, & Loeb, 2012; Miller, 2013). For example, Béteille et al. (2012) finds that the probability of turnover among higher value-added teachers is higher when the schools have a new principal, when compared to the probability among lower value-added teachers. They also find that principal turnover is associated with lower math achievement gains and suggest that this would be because new principals tend to hire less effective teachers, or they do not provide adequate professional support for new teachers. Ensuring sustainable leadership is, therefore, crucial for school reform and improvement (e.g., Hargreaves & Fink, 2004, 2006; Hargreaves, Moore, Fink, Brayman, & White, 2003).

Despite compelling anecdotes and the importance of stability and sustainability in school leadership, it remains unclear to what extent the NCLB sanction system is associated with principal turnover. Few empirical studies have examined this relationship. This study fills this gap in the literature by analyzing longitudinal administrative data and detailed school accountability data that I obtained from the Missouri Department of Elementary and Secondary Education (DESE). It uses distance to the AYP threshold as a key matching variable to identify comparable schools between those facing a sanction and those which were not. It provides new empirical evidence on the relationship between the NCLB sanction system and principal turnover.

Principal turnover is a two-sided event. From the principal side, as the anecdotes suggest, if they develop high levels of job stress and dissatisfaction due to accountability pressure, they may leave their schools when facing sanctions or sanction threats.[15] Turnover may be more likely to occur when the schools fail to make AYP for the first time, because the principals could perceive this first time failure to be a shocking event and an ominous warning for future

---

[15] In the remaining part of the paper, I use a term, NCLB sanctions, to incorporate both NCLB sanctions and sanction threats.

consequences. When departing from their schools, principals may transfer to different schools with lower probabilities of facing sanctions. Or, they could decide to take different roles at school buildings or district offices to avoid such sanctions. Furthermore, they may leave the public school system and work for private schools, non-profit organizations, or private firms.

From the district side, district administrators may replace principals who are working at schools that missed AYP with effective principals. The administrators could also assign them to high-performing or low-performing schools. Staff replacement and transfer is a standard practice among district administrators because principals' labor contracts are not as binding as teachers' and the principals often do not have tenure.[16] In Missouri, principals cannot acquire tenure in their positions (Missouri National Education Association, 2012).

This study uses reduced form models to estimate the association between the NCLB sanction system and principal turnover, and does not distinguish voluntary from involuntary turnover. I first investigate (1-a) to what extent the missing AYP for the first time is associated with principal turnover. As discussed earlier, principals may view a first-time failure as a shocking event, or a signal that they would face sanctions in the near future. Their responses to the first-time failure, therefore, may be distinctive. Missing AYP for the first time means that the principal faces informal sanction for the first time next year, which is public criticism and scrutiny over school activities. Principals may choose to transfer to different schools when they face informal sanction for the first time; they could take different jobs, such as district positions, assistant principalship, and teaching positions; they may also decide to leave the public education system entirely. So, I further investigate (1-b) whether the association is different by these turnover types. Moreover, principals' responses to the informal sanction could be different by

---

[16] According to the Schools and Staffing Survey 2011-12, 21 percent of the districts did not have a tenure system for the principals in 2011-12.

their qualifications, such as years of experience as a principal and the highest education level. Their responses may also vary by the characteristics of schools in which they work. For example, principals working at schools that enroll a large number of students eligible for the federal free/reduced lunch program may behave differently from their colleagues serving a small number of such students, because it might be challenging for the principals to intervene within a family setting to improve their lives and the students' performance. So I examine (1-c) whether the association is moderated by principal qualifications and student demographics.

Next, I shift the focus from the first time-informal sanction to all informal sanctions and actual NCLB sanctions. All informal sanctions include both first-time and non-first-time informal sanctions. I examine (2-a) to what extent the informal sanction and formal NCLB sanctions are associated with principal turnover. For formal NCLB sanctions, I focus on the first year of the NCLB sanction or School Improvement Year 1 (SIY 1), which required school districts to offer school choice options, because the number of schools facing sanctions beyond SIY 1 is too small to make reliable parameter estimates. For this part of the analysis, the sample includes only Title I schools because only those were subject to SIY 1. For the later sanction stages, I descriptively compare turnover patterns among Title I school principals who faced the sanctions with those not facing them. Like the first set of research questions, I also investigate (2-b) whether the association is different by turnover types and (2-c) whether the association is moderated by principal qualifications and school characteristics.

Finally, I descriptively examine (3-a) changes in school characteristics before and after principal transfer, and (3-b) the position the principals took when they faced informal sanction. Principals facing informal sanction may systematically choose to transfer to different schools with certain characteristics. They may also tend to take up certain positions. For these questions,

65

I focus on informal sanction, including first-time informal sanction, because the number of principals who transferred or took up different positions is too small in other samples.

A simple comparison of turnover behaviors between principals facing NCLB sanctions and those not facing them may lead to biased estimates, because schools facing such sanctions could be systematically different from those not facing them in both observable and unobservable ways. To reduce the potential bias from this selection, I utilize unique school-level assessment and AYP data and create a matching variable that measures distance to the AYP threshold in order to identify principals working at statistically similar schools except NCLB sanction status. This matching variable plays a critical role in reducing, if not removing, potential bias that results from the selection of schools into NCLB sanctions. I further adjust for remaining differences in observable principal and school characteristics through post-matching regression analysis.

Results from this study may help policymakers at different levels of government redesign their accountability systems. In general, incentives work only when policymakers design incentive structures appropriately. Findings from this study could help them identify the problems in the design, and what changes are needed to improve the effectiveness of accountability incentives. Policymakers may also use this study to design other incentive policies, such as performance pay and retention bonus programs.

This study proceeds as follows. The next section overviews the NCLB sanction system and specific features of AYP decision rules that the Missouri DESE applied. After the section, I discuss a conceptual framework and review the studies on NCLB and principal turnover. Then, I describe data and methods used to answer my research questions, after which, I report the results. I conclude with the implications and limitations.

NCLB Sanction System

General Description

NCLB required all public schools to make adequate yearly progress (AYP) in reading

and math from grades 3 to 8 and once at the high school level for the all-student group and all

student subgroups that meet a minimum subject-by-subgroup cell size requirement. [17] It also

required the schools to meet one additional academic indicator requirement, such as attendance

rate for elementary and middle schools and graduate rates for high schools. [18] Annual measurable

objectives (AMOs) for AYP were determined by each state education agency and they were

based on the percent of students performing at or above the state-defined proficient level on state

assessments.[19, 20] As the law gave states great discretion over the definition of proficiency

standards and AMOs, wide variation existed across them.[21] Student subgroups were typically

based on race and ethnicity, eligibility for the federal free/reduced lunch program, English

proficiency status, and disability status. In addition to the academic performance requirements,

the law required schools to meet a 95 percent participation rate requirement for all student

groups. If a school missed any of these requirements in any subject-by-subgroup group, it failed

to make AYP. Missing AYP in the same subject or in the same academic indicator for two

---

[17] All the information is based on the US Department of Education's (USDOE) desktop reference for NCLB (USDOE, 2002) unless indicated otherwise.

[18] The USDOE allowed flexibility in these requirements. While many states used reading assessments, a small number of states combined reading assessments with writing and/or language art assessments. The USDOE required the additional academic indicator from the 2005-06 school year. The USDOE gave the states discretion over what indicators to use. Many states used attendance rates for elementary and middle schools and graduation rates for high schools. A few states used writing or science assessment results or performance improvement measures as additional academic indicators. The cell size requirement also varied from state to state.

[19] A small number of states used a kind of performance index that combined the percent of students at each performance level (e.g., below basic, basic, meet the standards, exceed the standards) with larger weights assigned to higher performance levels.

[20] The law initially required AMOs to be raised to 100 percent by 2013-14. As of February, 2015, 43 states were granted waivers on the NCLB requirements, including the 100-percent proficiency requirement.

[21] To understand the achievement levels of students across the country, Banderia de Mello (2011) compared the proficiency standards of the National Assessment of Educational Progress (NAEP) with those of states' assessments. He found that there existed wide variation in the proficiency standards across the states, and that most states' proficiency standards are at or below NAEP's basic performance standards in both math and reading.

consecutive years placed the schools in the first year of the NCLB sanctions. It is important to note that only Title I schools were subject to these federal sanctions. Non-Title I schools were only required to make school improvement plans.

The NCLB sanction system was progressive and cumulative. Table 1 presents a summary of the sanction system. Although the first sanction started after missing AYP for two consecutive years, schools faced *informal* sanction when they missed AYP once, whether they were Title I schools or not. The informal sanction imply that schools were exposed to public criticism and scrutiny over school activities because state education agencies released school report cards to the public every year. The first formal sanction, School Improvement Year 1 (SIY 1), required school districts operating failing Title I schools to offer school choice options to eligible students enrolled in them with free transportation costs. The districts also had to provide technical assistance for the schools to help them improve their student performance. The second year sanction, SIY 2, required the districts to offer supplemental educational services (SESs) to the eligible students, in addition to school choice options.

Sanctions became more severe from the third year onwards. In this sanction stage, districts were required to take at least one corrective action, which included staff replacement and internal reorganization of the failing schools, while continuing to offer technical assistance, school choice, and SESs. The fourth and fifth year sanctions required districts to make school restructuring plans (SIY 4) and implement them (SIY 5) along with the requirements of the first three years of sanctions. Restructuring options included staff replacement, reopening as a charter school, school closure, state takeover, and contracting with private management companies for school management among others. Schools could exit sanctions only when they made AYP for two consecutive years.

The sanction system allowed the schools missing AYP to still make it through alternative routes. A first route was through the confidence interval, which was applied to either schools' actual performance (e.g., percent of students performing at or above the proficient level) or AMOs.[22] A second route was through the safe harbor, in which the schools could make AYP if they reduced the percent of students below the proficient level by 10 percent from the previous year. Some states combined the confidence interval with safe harbor.[23] Another route was through student growth models, which the USDOE first approved of in 2005-06. States were allowed to determine whether individual students could be classified as "on-track" to meet their individual growth targets. If so, the states could add these "on-track" students to the total number of students performing at or above the proficient level. Finally, schools could make AYP if the average of their last two to three years of performance exceeded AMOs in the current year. Not all states used all of these alternative routes.

Implementation of the NCLB Sanction System in Missouri

The Missouri DESE used Missouri Assessment Program (MAP), MAP-Alternative Assessment (MAP-A), and End-of-Course (EOC) assessments for AYP determination (Missouri DESE, 2010).[24] It tested students in grades 4, 8, and 11 in communication arts, and those in grades 3, 7, and 10 in math between 2001-02 and 2004-05. In 2005-06, DESE expanded the testing grades and started testing students in grades 3-8, and once at high school, in both subjects. DESE set a proficiency cut score for each grade in each subject, on which its annual school-level

---

[22] Most states used the confidence level between 95 percent and 99 percent.
[23] The confidence level ranged from 75 percent to 99 percent.
[24] The state changed their assessment in the 2004-05 school year.

proficiency target for every subject was based. Table 2 reports the annual measurable objectives from 2001-02 to 2013-14 in communication arts and math.[25]

The state required a minimum of 30 students in a subject-by-subgroup cell in order to determine AYP. This cell-size requirement was 50 for students with disabilities and English language learners until 2006-07. For the all-student group, if the number of students tested was less than 30, and the school did not meet AYP in a given year, the state aggregated the number for the most recent three years in order to determine AYP (i.e., uniform averaging procedure).[26] The state determined AYP, even if the number was still less than 30 after the three-year aggregation.

DESE used the 99-percent confidence interval, the safe harbor, the safe harbor with the 75 percent confidence interval, and student growth models to determine AYP.[27] The confidence interval was applied to the actual percent of proficiency achieved, and not the annual target. It was also applied to both the all-student group and student subgroups, whereas the safe harbor was applied to student subgroups only. Neither alternative route was applied unless the participation rate and cell size requirements were met with. DESE was approved to use student growth models in 2007-08 (Missouri DESE, 2010).

It is important to note that DESE did not have its own sanction system that was independent of the NCLB sanction system. DESE used a single accountability system and followed the NCLB system. No schools, whether Title I schools or otherwise, received additional state-level sanctions.

---

[25] Unlike most of the states, the Missouri DESE started making AYP determinations in 2001-02.
[26] For years from 2001-02 to 2003-04, the state used assessment data which were prior to the implementation of NCLB to determine AYP.
[27] DESE started using both the confidence interval and the safe harbor in 2003-04.

Conceptual Framework

Principal turnover results from many factors. For instance, principals may leave their schools because their salaries are not competitive enough, or they do not compensate for the tough job demands or poor working conditions. They may transfer to other schools because the accountability pressure in their current schools is so high that it increases their job stress and dissatisfaction levels. Or, from the demand side, district administrators may relocate them to new schools. To disentangle factors affecting principal turnover and simplify the analysis of the relationship between the NCLB sanctions and principal turnover, this study utilizes a labor market framework that is similar to one that Grissom and Andersen (2012) applied from the study of administrative turnover to study superintendent turnover.

In this framework, both principals and district administrators are viewed as actors in the principal labor markets, and principal turnover results from a two-sided decision. From the principals' perspectives, they make turnover decisions by evaluating their current jobs and comparing them with the alternative jobs available to them, including principal jobs at different schools, jobs at district central offices, and jobs outside the field of education. They calculate net benefits (or costs) of staying in the current schools and compare them with those of leaving those schools. If net benefits of leaving schools exceed the net benefits of staying put, principals turn over. Similarly, from the district administrators' perspectives, they determine whether to retain principals at their current schools, reassign them to different schools, or terminate their job contracts with them by calculating the net benefits of each option. If net benefits of one option exceed those of the others, district administrators would choose to take the most beneficial

option to them. These cost-benefit analyses are sometimes called the profitability test (Beach, 1990, as cited in Mitchell & Lee, 2001).

In the current context, principals working at schools and facing NCLB sanctions calculate costs and benefits of staying put and transferring. Costs may include an increasing amount of job stress and dissatisfaction due to NCLB sanctions, which may later cause health issues, such as cardiovascular disease (Johnson & Hall, 1988; Karasek, 1979; Schnall, Landsbergis, & Baker, 1994; Van der Doef & Maes, 1999). Benefits of staying could include relationships with colleagues, ongoing educational projects, and programs to which they are committed, and the community organizations with which they work.[28] At the same time, they calculate the costs and benefits of alternative jobs, such as principal jobs at better schools or private schools, administrative positions at district offices, and non-educational jobs. If an alternative job that is available to a principal is a principalship job at a high-performing school, benefits may include reduced job stress and dissatisfaction levels, less accountability pressure, talented teachers, among others. Costs may include a loss of relationships with one's current colleagues, students, parents, and community organizations. If the net benefits of staying are greater than the net benefits of leaving, these principals stay. Otherwise, they leave their schools for new positions.

Situating principals and district administrators in the principal labor markets also raises a need to consider their personal and professional characteristics, the characteristics of the schools they serve, and the characteristics of school districts in which they work, because these factors also play an important role in principal turnover. Prior studies find that principals' age, race/ethnicity, gender, professional experience, education level, and salary are associated with principal turnover (Akiba & Reicardt, 2004; Baker, Punswick, & Belt., 2010; Cullen & Mazzeo, 2008; Farley-Ripple, Solano, & McDuffie, 2012; Fuller & Young, 2009; Gates, Ringel,

___

[28] These are sometimes called job embeddedness (Mitchell & Lee, 2001).

Santibañez, Guarino, Ghosh-Dastidar, & Brown, 2006; Loeb, Kalogrides, & Horng, 2010; Solano, McDuffie, Farley-Ripple, & Bruton, 2010). They also suggest that student demographics, school enrollment size, school level, and urbanicity are all correlated with principal turnover. The number of schools in a district is also an important factor, because a larger district would have more job openings than the smaller ones (Farley-Ripple et al., 2012).

Figure 1 summarizes the conceptual framework. Although the two-sided nature of the turnover suggests the use of structural models, I use reduced-form models to estimate the relationship between the NCLB sanctions and principal turnover, which do not differentiate between decisions made by the principals from decisions made by the district. As a result, estimates on the influence of NCLB sanctions on principal turnover could be driven by either. Later, I will descriptively explore whether principal turnover was driven by principals or they were initiated by district administrators through examining the characteristics of the new hires.

NCLB and Principal Turnover

As briefly discussed earlier, the research base on school principals' responses to the NCLB sanction system has been thin. Among the limited number of studies on NCLB and principal turnover, Li (2012) has recently examined whether NCLB affected school principals' mobility by using longitudinal administrative data from North Carolina. She finds that NCLB lowered the average quality of school principals measured by value-added scores among the schools serving students from disadvantaged backgrounds. Principals with higher value-added scores tend to transfer to the schools that are less likely to face NCLB sanctions. In other words, NCLB shifted the distribution of principal quality across schools by affecting the principal mobility and hiring patterns.

Another recent study by Ahn and Vigdor (2014) has examined the relationship between school restructuring under the NCLB sanction system and principal turnover by using similar data from North Carolina, though the focus of the study is on student outcomes. They define principal turnover as an event where the principals in year *t* are not observed in any public school in the state in year *t+1*. This means that their analysis focuses on principal exits, rather than whether they transfer to different schools or change their positions. They find that school restructuring raised the probability of principal exit by 6 to 18 percentage points, although the impact was estimated imprecisely. White and Agarwal (2011) have descriptively analyzed principal turnover in Illinois and they have reported that even after controlling for confounding factors, schools that failed AYP experienced higher principal turnover rates.

Although not directly related to the NCLB sanction system, some studies on state and local accountability systems provide additional insights about principal turnover patterns. For example, Cullen and Mazzeo (2008) have examined principal turnover by using data from the state of Texas. Similar to the NCLB sanction system, the state's accountability system imposes sanctions, such as reconstitution, school closure, and private management (Texas Association of School Boards, 2008). They find that schools rated as low performing are 16.3 percentage points more likely to change their principals next year when compared to schools that are rated as exemplary. Schools rated as acceptable are 4.2 percentage points more likely to change their principals. Similarly, another study that examines the distribution of principal quality in Miami-Dade County Public Schools finds that principals at the schools receiving D or F in Florida's accountability system are more likely to leave them (Loeb et al., 2010).

These studies generally suggest that principals may be more likely to leave schools when they fail to meet AYP or face NCLB sanctions. They could transfer to different schools with

74

lower probabilities of facing NCLB sanctions, or they can exit the system. Yet, there are other important questions that these studies do not answer. For example, principals' responses may be different between first-time failure and non-first time failure. First-time failure could be a shocking event to the principals because it literally informs them that their schools performed below the state performance targets for the first time, and that the sanctions are just around the corner. As a result, principals might behave somewhat differently when their schools missed AYP for the first time. Although the focus is on student achievement, two studies find that first-time failure is associated with the gains in student test scores (Chakrabarti, 2014; Hemelt, 2011). These studies suggest that the effect of first-time failure could be different from that of non-first time failure.

Principals' responses to NCLB sanctions could be also different, as per their professional qualifications. For example, experienced principals may possess better skills to cope with accountability pressure and make effective decisions to improve student performance, whereas novice and inexperienced principals could struggle with the pressure and/or conflict with NCLB requirements, resulting in a decline in the AYP targets met (Dhuey & Smith, 2013; Shipps, 2012). In fact, principal experience is positively associated with school performance and negatively with turnover rates (Baker et al., 2010; Clark, Martorell, & Rockoff, 2009; Coelli & Green, 2012; Eberts & Stone, 1988; Fairley-Ripple et al., 2012; Solano et al., 2010). This may suggest some moderation effects through principal experience. Similarly, principals with advanced degree or those who attended competitive undergraduate institutions may have different skill sets to deal with accountability pressure, such as hiring teachers with stronger qualifications (Baker & Cooper, 2005) and stay with their schools even if they face NCLB sanctions.

Moreover, principals' turnover behaviors could be also different through the characteristics of schools they serve. For instance, as Title I schools are subject to formal NCLB sanctions, principals serving Title I schools may feel stronger accountability pressure than their colleagues serving non-Title I schools, resulting in higher turnover rates among Title I school principals when they face NCLB sanctions. School level may also matter, because more grades are tested at the elementary and middle school levels when compared to high school level, which could give elementary and middle school principals more challenges to meet annual performance targets at all tested grades. For student demographics, principals serving diverse, multicultural student populations typically face more difficult educational tasks than their colleagues serving predominantly white schools, in order to improve student performance (Gardiner & Enomoto, 2006; Jencks & Phillips, 1998). For example, principals at culturally diverse schools would have to deal with language issues (Brooks, Adams, & Morita-Mullaney, 2010; Gardiner, Canfield-Davis, & Anderson, 2009), work with parents with limited English proficiency and/or different perspectives on education and parenting, in order to engage them with school activities (Gardiner et al., 2009), and support students with disruptive home lives (Gardiner & Enomoto, 2005). Principals serving a large number of low-income students face more difficulties in while raising their performance levels (e.g., Rumberger & Palardy, 2005; Sanders, 1999). Poor families engage in children's education less, (e.g., Gutman & McLoyd, 2000; Lee & Bowen, 2006), teachers at high-poverty schools tend to have low expectations for their students (e.g., Boser, Wilhelm, & Hanna, 2014; Diamond, Randolph, & Spillane, 2004), and low-income students often have mental and health issues (e.g., Yoshikawa, Aber, & Beardslee, 2012). All of these could increase the accountability pressure and influence principals' responses to it.

Conclusively, this review of literature suggests that it may be equally important to differentiate first-time failure and investigate certain moderation effects through principal qualifications and school characteristics along with examining the relationship between NCLB sanctions and principal turnover. My study answers these research questions by using a unique distance variable to the AYP threshold and, subsequently, provides new empirical evidence.

Data

This study relies on multiple data sources to investigate the research questions in the state of Missouri. The Missouri Department of Elementary and Secondary Education (DESE) operates 523 regular public school districts and 2,430 public schools that enroll approximately 913,000 students. It is a relatively rural state, with 40 percent of the schools located in rural areas, compared with 34 percent in a median state. Missouri is, however, nationally representative in other important school characteristics. It serves 22 percent black or Hispanic students (with a median state being 25 percent) and 45 percent students who are eligible for the federal free/reduced lunch program (with a median state being 48 percent).[29] Student performance in the state is close to that of a median state. The average student test score for the state on the National Assessment of Educational Progress is within two points of the national median on fourth and eighth grade math and reading tests.

The primary data source is a longitudinal administrative data file obtained from DESE, which includes all personnel employed in the state education system between 1993-94 and 2009-10. Although all the analyses but one in this study are based on the data from 2001-02 to 2009-10, I use data prior to 2001-02 to construct a variable in order to measure years of principal

---

[29] These numbers and percentages are based on the author's calculations using the Common Core of Data 2012-13(district and school levels) maintained by the National Center for Education Statistics.

experience. I also use data prior to 2001-02 to perform a falsification test, which I describe later. The data file includes information about age, gender, race/ethnicity, highest degree, job/position code, which allowed me to identify school principals, position full-time equivalency, years of experience as an educator in the state, name of the undergraduate institutions attended, and annual salary, which was converted to 2009 constant dollars by using consumer price index for all urban customers (CPI-U) from the Bureau of Labor Statistics and adjusted for regional differences in the cost of living by using the Comparable Wage Index.[30] I converted the adjusted annual salary into salary that is relative to the average salary in a labor market, where a principal's school is located after controlling for principal and school characteristics. Information on undergraduate institutions was matched to Barron's ratings of college selectivity, in order to measure the principals' academic qualifications.[31] I create a variable that indicates whether a principal attended a competitive undergraduate institution. The data file does not include information about years of experience as a principal or the length of tenure as a principal in current school. I create these variables for those I observed entering the principalship after 1992-93.

Using the administrative data file, I create two turnover variables. First, I create a binary turnover variable that indicates whether a principal returned to his/her school next year in the same position. Second, I create a categorical turnover variable that indicates (1) whether a principal returned to his/her school next year as a principal, (2) whether a principal transferred to a new school as a principal in the same district or in a different district, (3) whether a principal

---

[30] I converted annual salary into 2009 constant dollars because the last year of the data (i.e., 2009-10) is used only to determine principals' turnover status. I used the Comparable Wage Index (Taylor, Glander, & Fowler, 2007). It is available from http://bush.tamu.edu/research/faculty/Taylor_CWI/.

[31] I use Barron's ratings in 1993 because a majority of principals attended undergraduate institutions in the late 1980s and the 1990s.

changed his/her position within the same school or at a different school, and (4) whether a principal exited the public education system.

These administrative data were merged with the data on school characteristics obtained from the Common Core of Data (school-level) between 2001-02 and 2009-10 maintained by the National Center for Education Statistics (NCES). These data include school locale (i.e., urban, suburban, town, and rural), student demographics (i.e., percent of nonwhite students, and percent of students eligible for the federal free/reduced lunch program), school level, and school enrollment size. I calculated the number of schools per school district. In addition, I merged these data with labor market information in Missouri that was obtained from the Comparative Wage Index data file, maintained by the NCES.[32]

I obtained detailed school-level assessment data used for AYP determinations from DESE, the Barnard/Columbia No Child Left Behind Database, 2002-03 and 2003-04 (Reback, Rockoff, Schwartz, & Davidson, 2011), and the National Longitudinal School-level State Assessment Score Database (NLSLSASD) created by the American Institutes for Research. DESE's accountability data file includes data on the number of students accountable, the number of students tested, the number and percent of students at or above the proficient level, attendance rates, graduation rates, and AYP results by subjects and subgroups, from 2004-05 to 2009-10. Using subject-by-subgroup AYP results and the information on additional academic indicators, I create a binary variable for school-level AYP results from 2004-05 to 2006-07. For years from

---

[32] The data file defines labor markets based on "place-of-work" as defined by the Census Bureau (Taylor, Glander, & Fowler, 2007). Based on this variable, I divided the state into the following 12 labor market regions: Saint Joseph, Northeast, Kansas City region, West Central, Joplin/Springfield, South Central, Cape Girardeau, Saint Louis Region, Columbia/Jefferson City, Lake of the Ozarks, Kansas City, and Saint Louis City. I treat Kansas City and Saint Louis City as single-district labor market regions. Taylor updated the data file recently, which is available from http://bush.tamu.edu/research/faculty/Taylor_CWI/.

2007-08 to 2009-10, school-level AYP results are available from a school improvement status data file that I obtained from DESE.

The Barnard/Columbia No Child Left Behind Database includes data on the number and percent of students tested, the percent of students at or above the proficient level, and AYP results by subjects and subgroups in 2003-04. It also includes school-level AYP results. Although the database includes accountability data in 2002-03, it does not include the data on the number of student tested as per subjects and subgroups, which prevents me from creating a variable to measure a distance to making AYP (see the method section). Fortunately, the NLSLSASD includes detailed school-level assessment data by subjects and subgroups, which include the number of students tested. Thus, for 2002-03, I use their data and merge them with AYP data from the Barnard/Columbia database. Similarly, for 2001-02, I use the NLSLSASD data. AYP results for that year were directly downloaded for each school from DESE's AYP Grid website.[33]

I obtained a list of schools in school improvement status from DESE in 2008-09. For years from 2003-04, which was the first year that the schools faced NCLB sanctions, to 2007-08, I collected a list of schools from DESE's websites using the Internet Archive Wayback Machine website and the Consolidated State Performance Report published on the US Department of Education website. Appendix Table 1 summarizes the data sources for school-level assessment data.

---

[33] The website provides overall AYP results by subjects and additional academic indicators, through which I create a variable for school-level overall AYP results. See http://mcds.dese.mo.gov/guidedinquiry/AYP/AYP%20-%20Grid.aspx.

<center>**Methods**</center>

Construction of the Distance Variable

Identifying the association between NCLB sanctions and principal turnover is challenging, because schools that faced the sanctions could be different from those that did not face them, in both observable and unobservable ways. An important step is to identify comparable schools that did not face the sanctions, but were similar to schools that faced them. Standard regression-based techniques will bias the estimates to the extent that unobservable factors such as parental involvement, community support, and school culture are associated with both NCLB sanction status and principal turnover. One way to find comparable schools would be to measure the distance to make AYP. Theoretically, if two schools, one making AYP and the other not making it, are located in similar positions at the distribution of the distance to the AYP threshold and have similar observable school and principal characteristics, they would be statistically identical. The difference in AYP status between them would be due largely to AYP determination rules that DESE used. I use this variation to identify the effect of NCLB sanctions on principal turnover.

The idea of the distance variable comes from the recent studies which used regression discontinuity designs to estimate local average treatment effects of the NCLB and its sanctions (e.g., Ahn & Vigdor, 2014; Chakrabarti, 2014; Fruehwirth & Traczynski, 2013; Hemelt, 2011). In these studies, researchers used similar forms of distance variables as a forcing variable in their design. I modify a distance variable employed by Fruehwirth and Traczynski (2013) and use it as one of the matching covariates in a propensity score matching method, which, as I explain later, to identify the comparable schools.[34]

---

[34] Initially, I planned to implement a regression discontinuity design. However, the distance variable, which I will describe shortly, did not pass a density test (McCrary, 2008). While it is unlikely that the distance variable was

<center>81</center>

Following Fruehwirth and Traczynski (2013), I create a variable to measure the distance in the following way. First, I compute the number of students in each subject for each student subgroup, including the all-student group that had to perform at or above the proficient level for a school to make AYP, incorporating the Safe Harbor and Safe Harbor with the 75% Confidence Interval.

Let j denote student subgroups and k subjects. P_{jkst} is the percent of students in student subgroup j in school s in year t that performed at or above the proficient level in subject k. Since DESE applies the 99 percent Confidence Interval to P_{jkst}, let CIP_{jkst} be the upper bound of the 99-percent confidence interval for P_{jkst}. This is the maximum percent of proficiency that each subgroup achieved for each subject.[35]

Let T_{kt} be the AYP proficiency level in subject k in year t set by DESE. Let SF_{jkst} be the target based on the Safe Harbor, which was calculated by $P_{jkst-1} + 0.1 \times (1 - P_{jkst-1})$, and $SFCI_{jkst}$ be the lower bound of the 75-percent confidence interval for the target based on the Safe Harbor.[36] Then, a school s made AYP in year t for subgroup j in subject k if

$$CIP_{jkst} \geq \bar{P}_{jkst}, \qquad\qquad (1)$$

where

$$\bar{P}_{jkst} = min\{T_{kt}, SF_{jkst}, SFCI_{jkst}\}.$$

Since the student growth model did not change the annual performance targets but added the number of students classified as "on track" to the number of proficient students, it is not included in $\bar{P}_{jkst}$ but rather, incorporated into CIP_{jkst}. The above expression does not include

participation rates or an academic indicator, because missing targets for participation rates, attendance rates, and/or graduation rates is quite different from missing AYP in reading or math for a student subgroup. It would take a school more effort to increase the number of proficient students than to increase participation rates or the numbers/percentages in additional academic indicators. For this reason, Equation (1) does not perfectly predict a school's AYP status.

Using CIP_{jkst} and $\bar{P}_{jkst}$, I compute the number of students who performed at or above the proficient level for subgroup j in subject k in school s in year t, TP_{jkst}, as well as the number of students that had to perform at the proficiency level and above, $\overline{TP}_{jkst}$. Let N_{jkst} be the number of students tested. Then,

$$TP_{jkst} = CIP_{jkst} \times N_{jkst} \text{ and}$$

$$\overline{TP}_{jkst} = \bar{P}_{jkst} \times N_{jkst}. \tag{2}$$

As a second step, I first take the difference between TP_{jkst} and $\overline{TP}_{jkst}$. Then, if $TP_{jkst} \geq \overline{TP}_{jkst}$ in all subjects for all subgroups, I aggregate the differences (all positive) across subjects and subgroups for each school s in each year t. If $TP_{jkst} < \overline{TP}_{jkst}$ in any subjects in any subgroups, I aggregate the negative differences across the subjects and subgroups for each school s in each year t. Mathematically, this measure takes a similar form to the one used by Fruehwirth and Traczynski (2013):

$$Testpasses_{st} = \begin{cases} \sum_{jk}(TP_{jkst} - \overline{TP}_{jkst}) & if \ TP_{jkst} \geq \overline{TP}_{jkst}, \forall j, k \\ \sum_{jk} 1\{TP_{jkst} < \overline{TP}_{jkst}\} \ (TP_{jkst} - \overline{TP}_{jkst}) & otherwise \end{cases} \tag{3}$$

where $1\{TP_{jkst} < \overline{TP}_{jkst}\}$ is an indicator variable that takes a value of 1 if $TP_{jkst} < \overline{TP}_{jkst}$. As a final step, I convert this measurement into a percentage measure, $\%Testpasses_{st}$, by dividing it by the total number of students tested in each subject-by-subgroup cell which exceeds the

minimum cell size requirements set by DESE.[37] I use this percentage as the distance to make

AYP. Theoretically, when a school makes AYP, $\%Testpasses_{st}$ is 0 or positive; otherwise, it is

negative. In practice, however, there are some false positives (i.e., missing AYP but the variable

is positive), because the distance measure does not incorporate other AYP requirements, such as

participation rates, attendance rates, and/or graduation rates. There are also some false negatives

(i.e., making AYP but the variable is negative) because of the appeal processes that might have

been in process while data were collected, because the school-level assessment data from the

NLSLSASD were not collected for AYP purposes, and/or because the data from DESE have

measurement errors.

Propensity Score Matching Procedures

The distance variable plays a critical role in identifying the comparable schools except

AYP or sanction status. However, it alone does not allow me to do so, because these schools

could be still quite different in other dimensions. To improve the matching quality, I use a

propensity score matching (PSM) method. This matching is based on year *t* or a baseline year

because the actual treatment is given next year (year *t+1*). Figure 2 shows when the matching is

performed and when the treatment (i.e., sanction) is given.

First, I estimate the probability of missing AYP in a baseline year or the propensity

scores as a function of characteristics of principals, schools, and districts, through standard

logistic regression techniques. More formally, I estimate the following logistic regression model:

$$\Pr(missing\ AYP) = \frac{e^{f}}{1+e^{f}}, \tag{4}$$

---

[37] Students are counted more than once if they belong to multiple subgroups.

where

$$f = \beta_0 + P_{isdt}\beta_1 + S_{sdt}\beta_2 + D_{dt}\beta_3 + \varepsilon_{isdt}.$$

The probability that a school s where a principal i works misses AYP in a baseline year t is a function of principal characteristics P_{isdt} (i.e., age, female, race/ethnicity, selectivity of undergraduate institutions that a principal attended, holding an education specialist or doctorate degree, years of principal experience, years in current school, years in the Missouri education system, and relative salary), school characteristics S_{sdt} (i.e., distance to making AYP, urbanicity, school level, school enrollment size, percent of nonwhite students, percent of students eligible for the federal free/reduced lunch program, and Title I school status), district characteristics D_{dt} (i.e., number of schools), and a random error term ε_{isdt}. Principal characteristics may not be directly associated with the schools' AYP status. Yet, as they affect turnover or the outcome variable, I include them to increase the precision of the estimates (Brookhart, Schneeweiss, Rothman, Glynn, Avorn, & Stürmer, 2006; Stuart, 2010). The baseline year varies from principal to principal because it is the year when the school missed AYP. The last year of the baseline year is 2007-08 because the treatment year becomes 2008-09 in this case, and I need administrative data for 2009-10, which is the last year of the data, to identify whether a principal turned over at the end of 2008-09.

After estimating the propensity score for each principal, I perform one-to-five nearest neighbor matching within the common support region with replacement.[38,39] Following a recommendation from the literature, I set a caliper width at a quarter (0.25) of the standard

---

[38] I also performed one-to-one nearest neighbor matching and radius matching with a caliper width restriction, but the matching quality was the best for one-to-five nearest neighbor matching in terms of standardized bias and the number of units that lie within the common support region.

[39] Due to the distance variable, about 13 percent of principals in the treatment group are outside of the common support region in the sample of principals whose schools had not missed AYP in the baseline year. About one percent of the principals lie outside the common support region for the sample to examine the first part of the second research question (i.e., informal sanction); about 19 percent of the principals are outside the region for the sample to investigate the second part of the research question (i.e., School Improvement Year 1).

deviation of the logit of the propensity scores (Austin, 2011; Rosenbaum & Rubin, 1985). I assess the matching quality based on standardized bias (or difference) in the covariates between the treatment and comparison groups and use 25 percent or below as a criterion to determine whether the covariates are balanced (Harder, Stuart, & Anthony, 2010; Rubin, 2001; Stuart, 2010).[40] If any one of the covariates exceeds this threshold, I include a combination of various forms of variables (e.g., quadratic and cubic forms) and/or the interaction terms among the covariates exceeding the threshold until all standardized bias becomes below the threshold.

Based on this set of matched schools, I estimate the probability of principal turnover at the end of the treatment year, Year $t+1$, using an indicator variable for sanction status and the same set of the covariates used in Equation (4), except that I drop the distance variable to avoid the issue of statistical overcontrol and that year fixed effects and labor market region fixed effects are added to the equation.[41] The model takes the following form:

$$\Pr(Turnover)_{isdt+1} = \frac{e^f}{1+e^f}, \tag{5}$$

where

$$f = \beta_0 + Sanction_{sdt+1}\beta_1 + X_{isdt}\beta_2 + S_{sdt}\beta_3 + D_{dt}\beta_4 + \gamma_t + \pi_t + \varepsilon_{isdt+1}.$$

The probability that a principal i in school s in district d in the treatment year, year $t+1$, turns over at the end of year $t+1$ is a function of NCLB sanctions $Sanction_{sdt+1}$, baseline principal and school characteristics X_{isdt}, S_{sdt}, year fixed effects γ_t, labor market region fixed effects π_t, and a random error term ε_{isdt+1}. If the coefficient on the sanction variable is equal to or greater than 1, that indicates that principals are more likely to leave their schools when they face sanctions.

Next, I use a categorical turnover variable to investigate whether NCLB sanctions are associated with certain types of turnover behaviors (i.e., movers, position changers, and leavers)

---

[40] The standardized bias (or difference) is the mean difference as a percentage of the average standard deviation. See Rosenbaum and Rubin (1985) for the formula to calculate standardized bias.

[41] As most of the principals were matched across districts, district fixed effects are not used.

by estimating a series of multinomial regression models with staying put as a base category.[42] This model uses the same covariates used in Equation (5) and takes the following form:

$$\Pr(\textit{turnover type} = j)_{isdt+1} = \frac{e^{f_j}}{1 + \sum_{k=2}^{4} e^{f_k}}, \text{ for } j > 1 \tag{6}$$

where

$$f_j = \beta_{j0} + \textit{Sanction}_{sdt+1}\beta_{j1} + X_{isdt}\beta_{j2} + S_{sdt}\beta_{j3} + D_{dt}\beta_{j4} + \gamma_{jt} + \pi_{jt} + \varepsilon_{jisdt+1}.$$

For my entire analysis, I exclude those principals who did not stay in the same schools in the baseline year and the treatment year, because the propensity score matching is based on the baseline year. To answer the first set of research questions about the association between the first time informal sanction and principal turnover (RQs 1-a, 1-b, and 1-c), I first restrict the sample to the principals working at schools that missed AYP for the first time and those working at schools that had never missed AYP in the baseline year. From this sample, I perform the PSM described above. The first group of principals faces the informal sanction for the first time in the treatment year; the second group does not. For the second set of research questions (RQs 2-a, 2-b, and 2-c), I first restrict the sample to those working at schools in a given sanction stage (i.e., informal sanction and SIY 1) in the baseline year. Then, I perform the matching. Again, the treatment group faces the informal sanction or SIY 1 in the treatment year; the control group does not. For the informal sanction, the sample includes those working at Title I and non-Title I schools. From School Improvement Year 1 (SIY 1), the sample is restricted to the principals working at Title I schools only. As explained earlier, I descriptively analyze the patterns in principal turnover for SIY 2 and beyond.

---

[42] This variable takes the value 1 for stayers, 2 for movers, 3 for position changers, and 4 for leavers.

Results

First Time Informal Sanction and Principal Turnover (Research Question 1)

In this section, I examine the relationship between first-time informal sanction and principal turnover. Missing AYP for the first time in the baseline year means that the principal faces informal sanction next year, which is the treatment. I first take a descriptive look at the pre-matching sample of principals who worked at schools that had not faced any sanction in the baseline year. Table 3 reports the descriptive statistics on characteristics of principals, schools, and districts by sanction status. The first two columns display the average characteristics of those principals not facing the informal sanction in the treatment year; the next two columns for those facing it. Their characteristics are compared by using a series of t tests. It shows that these two groups of principals are quite different in many characteristics, especially their school characteristics. For example, close to 50 percent of principals facing the sanction are female, whereas just over 60 percent are female among their colleagues who are not facing the sanction. By race, 12 percent of principals facing the sanction are black; and only four percent of principals not facing it are black.

For school characteristics, there is a large gap in the distance variable between the two groups (4 percent and 29 percent, respectively). The average percent of students eligible for the federal free/reduced lunch program is approximately 45 percent among principals facing the sanction next year, whereas it is 37 percent among their colleagues who are not facing it. By school level, 47 percent of the principals facing the sanction work at elementary schools, but over 80 percent of their colleagues not facing it serve elementary schools. There is also a notable difference in the number of schools per district. The average number of schools is about 22 among those facing the sanction, whereas it is 15 among those not facing it. These differences

highlight the importance of the use of propensity score matching in order to balance the differences between these two groups of principals.

Table 4 reports principal turnover rates by turnover types and the treatment status in the pre-matching sample. Stayers are those principals who returned to schools as principals. Movers are those who transferred to new schools as principals. Position changers are those who changed their positions and worked at the same schools or different schools. Finally, leavers are those who exited the state public education system. Again, turnover rates are compared by the sanction status using a series of *t* tests.

About 19 percent of principals whose schools had never missed AYP left their schools, whereas about 25 of them under the informal sanction departed from their schools. By turnover types, principals facing the informal sanction are more likely to change their positions or leave the public education system when compared to those facing no sanctions. Transfer rates are statistically indistinguishable between the two groups, whether within-district or across-district transfers.

As discussed in the method section, schools facing informal sanction could be different from those not facing any sanctions in both observable and unobservable ways. To balance the covariates between the treatment and comparison groups, I performed a propensity score matching described in the section and checked the covariate balance between the two groups. Table 5 reports results from the balance test. The first two columns show the means of the covariates for the comparison and treatment groups, respectively. The third column displays the standardized bias, which is the mean difference as a percent of the average standard deviation (Rosenbaum & Rubin, 1985). It falls within a range between -25 and 25, which suggests that the

propensity score matching achieved a balance in all of the covariates (Harder et al., 2010; Rubin, 2001; Stuart, 2010).

Now, I turn to estimation results. Table 6 reports results from logistic regression of the probability that a principal turns over at the end of the treatment year. The treatment is first-time informal sanction (i.e., public criticism and scrutiny). The comparison group includes those principals whose schools had never missed AYP, and thus, did not face the sanction in the treatment year. The first three models report results from logistic regression without the PSM; the next three models are estimated after the PSM. Models 2, 3, 5, and 6 include interaction terms between the sanction status variable and an indicator variable for Title I status or high school to test whether principals working at these schools behave differently. All models include characteristics of principals, schools, and districts, as well as year fixed effects and region fixed effects. Coefficients are transformed into odds ratios, and z statistics are reported in parentheses. For brevity, I only report odds ratios for the sanction status and the interaction term.

The table shows that facing the informal sanction for the first time is positively associated with higher turnover rates in a naïve logistic regression model (Model 1). In a matched sample, the association becomes statistically insignificant (Model 4), but the sign of the odd ratio is in the expected direction. I find no evidence that the association is moderated by Title I status or school level, whether PSM is used or not.

The association could be moderated by principals' qualifications or the characteristics of schools that they serve. To test these possibilities, I add interaction terms between the sanction status variable and each of the following principal and school characteristic variables separately, and then estimate the models after PSM: attended a selective undergraduate college, holding an education specialist degree or a doctoral degree, total years of principal experience, percent of

nonwhite students, and percent of students eligible for the federal free/reduced lunch (FRL) program. To test for non-linearity in moderation effects, I also create three indicator variables for years of principal experience and student demographics. I replace each of the continuous variables with these three indicator variables separately and interact each of them with the sanction status variable. Table 7 displays the results. Panel A shows results for principal qualifications; Panel B reports results for student demographics.

I find little evidence that the relationship between the sanction status and principal turnover is moderated by principal qualifications or student demographics. Although estimated imprecisely and thus suggestive, the relationship appears to be somewhat moderated by years of principal experience. For example, the odds ratio on the interaction term between the sanction status and years of principal experience is 0.92 ($p=0.170$, Panel A, Model 3). This suggests that the difference in the predicted probability by sanction status (i.e., facing the sanction minus not facing it) becomes smaller, or even turns into negative, as the principals accumulate years of experience. In other words, it suggests that the difference is positive and widest among the least experienced principals.

Model 4 tests for non-linearity in this relationship. To ease the interpretation of the results, I plot the predicted probability of turnover by sanction status over the four principal experience categories in Figure 3. The predicted probability of turnover among the least experienced principals is 20 percent when they faced informal sanction for the first time; it is only 13 percent when they did not face informal sanction. Although the difference of seven percentage points appears to be large, it is estimated imprecisely ($p=0.117$). At the other end of the spectrum, the predicted probability of turnover among the most experience principals is 25 percent when they faced informal sanction for the first time; it is 35 percent when they did not.

The difference turns from positive into negative, and the difference in these differences is close to 18 percentage points and statistically significant (p=0.061).

My analysis in Tables 6 and 7 combines all types of turnover (i.e., movers, position changers, and leavers) and examines a binary turnover variable. Yet, it is possible that informal sanction may be positively associated with a certain type of turnover behaviors, as Table 4 suggests. To investigate this possibility, I divide turnover into (1) transfers to new schools, (2) position changes, and (3) exits, and estimate the multinomial logistic regression models (Equation (6)) with staying in current position as the base category. The models include the same set of covariates used as before.[43] Table 8 displays results based on the matched sample. Again, for brevity, I only report the relative risk ratios for the sanction variable and interaction terms. The table shows no evidence that the informal sanction is associated with any types of turnover, whether a principal works at a Title I school or a high school.

Like the previous analysis, I also examine whether the relationship for each turnover type is moderated by principal qualifications and student demographics. As I do not observe any notable moderation effects by student demographics, I focus on the moderation effects by principal qualifications. Panel A of Table 9 reports results for college selectivity and highest degree; Panel B shows results for years of principal experience.

In Panel A, I find no evidence that college selectivity moderates the association for any turnover types (Model 1). For highest degree, although estimated imprecisely and thus suggestive, the relationship between informal sanction and position changes appears to be negatively moderated by degree status, suggesting that principals without an education specialist or a doctoral degree are more likely to change their positions when they face the informal sanction for the first time (Model 2). The difference in the probability of position changes among

---

[43] A variable for other race was dropped due to the convergence problem.

principals without an education specialist degree or a doctoral degree is about three percentage points (p=0.112). The difference becomes negative two percentage points among those with an education specialist or a doctoral degree, and the difference in these differences is about five percentage points (p=0.144).

In Panel B, Model 3 shows strong evidence that the relationship between the sanction status and the probability of position changes is moderated by years of principal experience. This suggests that first-time informal sanction affects inexperienced principals more strongly than experienced ones in their position change behaviors. Model 4 tests for non-linearity in moderation effects. I find clear evidence that moderation effects are non-linear for the probability of transfer and position changes. To ease the interpretation of the results, the predicted probabilities of transfers and position change by the sanction status are plotted over years of principal experience in Figures 4 and 5, respectively. In Figure 4, the predicted probability of transfer among principals in the first experience category (i.e., with two to three years of experience) is five percent if they face informal sanction for the first time; however, it is only two percent if they do not. This difference is statistically different from zero (p=0.061). The difference in the probability in the next two categories becomes negative but jumps to eight percentage points in the last experience category. The predicted probability of transfer is eight percent when facing informal sanction; it is less than one percent when not facing it. The figure shows that the first-time informal sanction affects the least and the most experienced principals more strongly than those in the middle experience categories, although the difference in the differences in the probabilities between the least and most experienced groups is statistically indistinguishable from zero (p=0.115).

Figure 5 shows that the predicted probability of position changes follows similar patterns, except in the last experience category, where the probability of position changes is higher among those principals not facing the informal sanction (Figure 5). The probability of position changes is 13 percent among the most experienced principals who are not facing the sanction; it is eight percent among those facing the sanction. Yet, like transfer rates, first-time informal sanction affects the least experienced principals more strongly in their position change behaviors. The probability is nine percent when inexperienced principals face the informal sanction; but it is only two percent when they do not. This difference is statistically significant.

Each NCLB Sanction and Principal Turnover (Research Question 2)

In this section, I examine the influence of the informal sanction, whether first time or not, and each stage of NCLB sanctions on principal turnover. I first investigate whether facing informal sanction or SIY 1 is associated with principal turnover. Then, I descriptively analyze the influence of later sanctions on principal turnover because a smaller number of schools faced SIY 2 and beyond, which prevents me from finding good matches.

Influence of the Informal Sanction/SIY 1 on Principal Turnover

For this research question, I examine two different samples. For informal sanction, my sample includes both Title I and non-Title I school principals because informal sanction is public criticism and scrutiny that both principals faced when their schools failed to make AYP. On the other hand, SIY 1 was a sanction that only Title I school principals were subject to. So, the sample for SIY 1 includes Title I school principals only. Descriptive statistics in the pre-matching samples and balance test results are presented in Appendix Tables 2 and 3.

94

First, I take a descriptive look at principal turnover rates. Table 10 displays turnover rates by sanction status in the pre-matching sample. The left side of the table reports turnover rates among those who did not face any sanction in the baseline year. I find very similar patterns found among the sample of principals whose schools had never missed AYP until the baseline ear (see Table 4). The average turnover rate among those facing informal sanction in the treatment year is about 25 percent, whereas it is about 20 percent among those not facing it in that year, a significant difference of five percent points. By turnover types, I find that the principals facing informal sanction are more likely to change their positions or leave the public education system. The right side shows turnover rates among Title I school principals who faced the informal sanction in the baseline year. I find weaker patterns. Title I school principals facing SIY 1 are less likely to transfer, but more likely to exit the system.

Now, I turn to estimation results. Panel A of Table 11 reports results for the association between the informal sanction and principal turnover. It shows that, although the main effect is insignificant, the relationship is moderated by Title I school status. The predicted probability of turnover among Title I school principals is 23 percent when they face informal sanction, whereas it is 16 percent when they do not. The sign of inequality flips among non-Title I school principals. The predicted probability is 27 percent when they do not face informal sanction; but it *decreases* to 24 percent when they face it. I do not observe a moderation effect by high school. Panel B displays results for the association between SIY 1 and turnover. I find no evidence that facing SIY 1 is associated with principal turnover rates.

Next, I examine the moderation effects by principal qualifications and student demographics. Like the previous analysis, because I do not find any notable moderation effects by college selectivity, the highest degree level, or percent of FRL students, I focus on the total

years of principal experience and the percent of nonwhite students. Table 12 reports results. In Panel A, I find strong evidence that the association between informal sanction and turnover rates is negatively moderated by years of principal experience and but positively moderated by the percent of nonwhite students.

To ease the interpretation of these results, I again plot the predicted probability of turnover which is based on Models 2 and 4 by sanction status over the four principal experience/percent nonwhite student categories in Figures 6 and 7, respectively. Figure 6 shows very similar patterns to those found in Figure 3, but differences in the predicted probability by sanction status are wider. Principals with two to three years of principal experience are about 13 percentage points more likely to leave their schools when they face informal sanction, compared to their colleagues with the same experience level but not facing the sanction (24% and 11%, respectively). In the last experience category, on the other hand, principals facing informal sanction are about 15 percentage points *less* likely to turn over than those not facing it (27% and 42%, respectively). As the table indicates, the difference in the differences between the two experience categories (i.e., about 28 percentage points) is statistically significant.

Figure 7 displays a notable moderation effect. The difference in the predicted probability of turnover between principals facing the informal sanction and those not facing it widens, as the percent of nonwhite students increases. At the lowest level, the difference is about negative two percentage points and statistically insignificant; the sign of the difference flips in the next experience level and the difference increases to about seven percentage points (19% and 12%, respectively). The difference jumps to about 17 percentage points in the next two levels. This suggests that informal sanction affects principals serving a larger number of nonwhite students more strongly than those serving a small number of such students. In Panel B, in contrast to the

informal sanction, I find little evidence that the association between SIY 1 and turnover is moderated by years of principal experience or the percent of nonwhite students.

Table 13 reports results from the same analysis by turnover type. I observe similar patterns. In Panel A, I find that Title I school status positively moderates the relationship between informal sanction and exit behaviors. The predicted probability of exit is 13 percent among Title I school principals facing informal sanction, whereas it is only five percent among their colleagues who are not facing it. The difference in the predicted probability among non-Title I school principals is statistically indistinguishable from zero. Although it is estimated imprecisely and thus suggestive, Title I status also positively moderates the relationship between informal sanction and transfer behaviors. Title I school principals are more likely to transfer to different schools than non-Title I school principals when they face informal sanction (p=0.166).

One difference from the binary turnover analysis is that high school principals are much more likely to leave the system when they face informal sanction. Among high school principals, the difference in the predicted probability of exit between those facing informal sanction and those not facing it is eight percentage points (10% and 2%, respectively); however, the difference is just about one percentage point among those working at the other school levels (12% and 11%, respectively). Here, the difference in the differences is about seven percentage points and statistically significant. On the other hand, SIY 1 is found uncorrelated with any turnover types.

Like the binary turnover analysis, I also examine whether principal qualifications and student demographics modify the relationship. Although results are not reported, I find similar patterns. Years of principal experience moderates the association between informal sanction and turnover by types, and the difference in the predicted probability of transfer or position changes by sanction status is much wider among principals with two to three years of experience. The

97

principals facing the sanction are much more likely to transfer or change their positions. On the

other hand, the moderation effect by the percent of nonwhite students found in the binary

analysis largely comes from their effect on principal transfers and exits. Principals serving a

large number of nonwhite students are much more likely to transfer to different schools or exit

the system when they face the informal sanction, compared with those in similar schools who are

not facing the sanction.

Influence of SIY 2 and Beyond on Principal Turnover

As discussed earlier, only a small number of schools faced SIY 2 and sanctions beyond

SIY 2 in the pre-matching sample, which prevents me from find good matches. Instead, I

descriptively explore the relationship between these sanctions and principal turnover. Table 14

displays principal turnover rates by sanction status among those facing different stages of

sanctions. I first restrict the sample to each sanction stage in the baseline year (i.e., SIY 1, SIY 2,

SIY 3, and SIY 4). Then, I divide each sample into two groups, based on their sanction status for

next year. For example, if a school that faced SIY 1 in the baseline year missed AYP at the end

of the year, it faces SIY 2 the next year. If it made AYP, it faces SIY 1 delayed, which means

that their sanction status is held for one year. Using these two groups, I compare turnover rates

through a series of *t* tests.

In the first sample (i.e., facing SIY 1 in the baseline year), the average turnover rate

among principals facing SIY 2 is 34 percent, whereas it is 24 percent among those facing SIY 1

delayed. In the next sample, the difference in the average turnover rate between those facing SIY

3 and SIY 2 delayed is about 19 percentage points. These differences are not statistically

indistinguishable from zero due to the small sample sizes. Larger sample sizes may turn them into being statistically significant. The last two samples are too small to interpret the results.

For the first sample, I also estimate a series of logistic regression models that are similar to Equation (5), without propensity score matching and test for moderation effects by principal qualification and student demographics. Although results not reported, I do not find evidence that SIY 2 is associated with principal turnover or that the relationship is moderated by principal qualifications or student demographics.

Falsification Test

All of the analyses performed for Research Questions 1 and 2 assume that the propensity score matching method and the subsequent post-matching logistic and multinomial logistic regression models account for bias due to potential endogeneity in the sanction status variable. However, it is still possible that the variable is correlated with other unknown, unobservable factors, also associated with turnover outcomes. For example, if these omitted variables are positively correlated with the sanction status variable and negatively correlated with the turnover variable, the estimates are biased downward and could be insignificant. If this is the case, the null results for the main analyses may be misleading.

To provide evidence on this possibility, I perform a falsification test, which examines whether the sanction status variable in post-NCLB years predicts principal turnover behaviors in the pre-NCLB years. Theoretically, NCLB sanction cannot predict turnover behaviors in the pre-NCLB years. However, if there are omitted variables that are positively correlated with the sanction status and negatively associated with the turnover variable, then the test would find that the sanction status is significantly, negatively associated with turnover outcomes in the pre-

NCLB years (i.e., more likely to stay). This suggests that the sanction status is endogenous and the main results are misleading. I conduct this test only for the main turnover analyses.

For this test, I use schools in each matched sample and predict principal turnover as a function of the sanction status in the post-NCLB years, characteristics of principals, schools, and districts in the pre-NCLB years from 1993-94 to 1999-2000, year fixed effects, and labor market region fixed effects.[44] Most of the principals in these years would not be the same as the ones in the post-NCLB years. Table 15 reports results from the test. It shows little evidence that the sanction status predicts the turnover outcomes in the pre-NCLB years, alleviating some concern that the sanction status variable is endogenous.

Analysis of Principal Transfers and Position Changes (Research Question 3)

My analysis, so far, focused on whether a sanction is associated with principal turnover. This subsection takes a closer look at the patterns in principal transfer and position changes. NCLB sanction may have influenced principals' transfer patterns and career trajectories in a systematic way.

Transfer Patterns

As discussed in the theory section, when principals decide whether to move to different schools, they perform profitability tests or cost-benefit analyses, by comparing current schools with future schools in terms of both unobservable and observable characteristics. Principals may prefer to work at schools with certain characteristics, particularly those with a lower probability

---

[44] Because data on students eligible for the federal free/reduced lunch program and Title I school status become available from 1998-99 in CCD, I dropped the Title I status variable from the models and replaced the percent of students eligible for the free/reduced lunch program with the percent of students eligible for the free lunch program. In addition, the weights generated from the propensity score matching are not used for this analysis.

of missing AYP or facing a sanction. While my empirical models cannot incorporate future job

characteristics, I can provide descriptive information about the principals' transfer behaviors by

examining the characteristics of "sending" schools and "receiving" schools that are potentially

associated with AYP results and testing whether the differences are statistically indistinguishable

from zero. I examine changes in the following school characteristics: distance to making AYP,

percent of students eligible for the free/reduced lunch program, percent of nonwhite students,

school enrollment size, Title I status, and school level. For this analysis, I focus on two groups of

principals who did not face any sanction in the baseline year: (1) principals who did not face

sanction in the treatment year, and (2) principals who faced informal sanction during that year.

The left side of Table 16 reports results for the first group; the right side reports results for the

second group.[45] . The last two columns show differences in the differences between the two

groups and associated p values. Panel A includes all principals; Panel B includes Title I school

principals. Because sample sizes are small, which yield relatively large p values, I do not discuss

the results from t tests.

I observe three important patterns. First, whether they face the informal sanction or not,

principals are more likely to transfer away from Title I schools. For instance, in Panel A, among

all principals who faced informal sanction and transferred at the end of the treatment year, 56

percent of them worked at Title I schools before transfer. The percent declines to 44 percent after

transfer. Similarly, in Panel B, among Title I school principals facing the sanction, 95 percent of

them worked at Title I schools before transfer[46]; it declines to 50 percent after transfer. It is clear

---

[45] I also examined changes in the school characteristics for those principals whose schools never missed AYP and those whose schools missed AYP for the first time in the baseline year, and thus, faced informal sanction in the treatment year. I find similar patterns with minor differences in significance due to small sample sizes. I did not perform this analysis for those facing SIY 1 or the later years of sanctions because the number of principals who transferred is too small.

[46] This is not 100 percent because Title I school status changed between the baseline year and the treatment year. Data on school characteristics among the sending schools came from the treatment year.

that principals tend to avoid Title I schools. Along with this move, principals tend to transfer to

schools that serve a smaller number of students who are eligible for the federal free/reduced

lunch program. On the other hand, the percent of nonwhite students does not change very much

before and after transfer. Third, principals do not appear to be concerned very much about the

distance to the AYP threshold. Changes in the distance variable are minimal, whether principals

face the informal sanction or not and whether they work at Title I schools or not. Overall, I do

not observe notable differences in the transfer patterns between those facing informal sanction

and those not facing it.

Patterns in Position Changes

Next, I examine what kind of positions principals took after they faced informal sanction

and whether the patterns in their position changes are systematically different from those who

did not face the sanction but changed their positions. As in the previous section, I focus on two

groups of principals who did not face any sanction in the baseline year: (1) principals facing

informal sanction in the treatment year, and (2) principals facing no sanction during that year.

Table 17 displays results by sanction status. A Chi-squared proportion test shows that there is no

significant relationship between the sanction status and types of positions that the principals took.

However, there are some notable patterns. First, principals tend to take positions at the district

central office, whether they face informal sanction or not. For instance, about 43 percent of

principals who did not face the sanction took central office jobs; about 40 percent did so among

those facing the sanction. Second, about one in five principals change their position's full time

equivalency from 0.75 or above to below 0.75. This means that their job duties were reduced

and/or that they were assigned to other jobs at the same time. Third, principals facing the

informal sanction tend to become assistant principals or classroom teachers, when compared with those not facing it (27% and 20%, respectively).

Distinguishing Voluntary and Involuntary Turnover

As described in the conceptual framework section, all turnover models are reduced form models, and they do not distinguish between voluntary and involuntary turnover. Yet, it is important to know whether turnover was initiated by principals or district administrators, because it affects the interpretation of the results. If turnover was initiated by the principals, my results may have policy implications for retaining them. On the other hand, if it was initiated by district administrators, my results are not sufficient to provide useful policy implications, unless I examine what happened to the schools after turnover. If I find that the school performance improved after replacing principals, I can conclude that the districts' decisions were effective. If it did not improve, then the decisions were not effective.

To test whether turnover was systematically initiated by the principals or not, I descriptively compare professional characteristics of those principals who left their schools and those of new hires. If turnover was initiated by district administrators, I would expect that principal qualifications systematically improved after turnover, because the main reason for principal replacement would be to bring in more effective principals to raise school and student performance. I examine the same set of principal qualifications that I used in the previous analysis. These include college selectivity, highest degree attained, and years of principal experience. Table 18 reports results. Panel A displays results for the first time informal sanction; Panel B for the informal sanction, including both first time and non-first time; and Panel C for SIY 1. The first set of three columns in each panel compares principal qualifications for schools

103

which did not face informal sanction or SIY 1. The second set compares the same qualifications for those schools facing them. The last column reports the differences in the differences and indicates whether the differences are statistically significant.

The table shows no evidence that the principal qualifications systematically improved after the turnover, whether facing the sanction or not. Rather, schools systematically hired principals with weaker qualifications. Furthermore, I find no evidence that schools facing the sanctions hired principals with stronger qualifications than schools not facing them. Although principal performance measures are not examined in this analysis, these results suggest that principal turnover might not be systematically initiated by district administrators. It is more likely that such turnover was largely initiated by the principals themselves.

Discussion and Conclusions

Despite the critical role that principals play in improving student performance and the importance of stability in school leadership for successful school improvement (e.g., Branch, Hanushek, & Rivkin, 2012; Brewer, 1993; Grissom, Loeb, & Master, 2013; Hallinger & Heck, 1998; Hargreaves & Fink, 2004; Hargreaves et al., 2003; Waters, Marzano, & McNulty, 2003), few studies have examined whether NCLB sanctions have influenced principals' turnover behaviors. This study fills this gap in the literature by using longitudinal administrative data and detailed school-level assessment data for AYP determinations from Missouri. It constructs a distance variable to the AYP threshold to identify statistically comparable schools between those facing the sanction and those not facing it through a propensity score matching method. It offers important key findings about principals' turnover behaviors under the NCLB sanction system.

First, the informal sanction is not statistically associated with higher turnover rates, whether facing the informal sanction is the first time or not. However, principals' responses to informal sanction are different by their job experience level and the percent of nonwhite students at the schools they serve. For example, principals with two to three years of principal experience are about 13 percentage points more likely to leave their schools when they face the informal sanction, including both first time and non-first time, when compared with their colleagues with the same experience level but not facing the sanction (24% and 11%, respectively). In contrast, experienced principals (i.e., 10 years of experience or more) facing the informal sanction are about 15 percentage points *less* likely to turn over when compared to those not facing it (27% and 42%, respectively). The difference in these differences (i.e., 13% and -15%) is statistically significant.

I also find a clearer moderation effect by the percent of nonwhite students. The difference in the predicted probability of turnover by the informal sanction status among principals serving a larger number of nonwhite students is much wider than the difference among principals serving a smaller number of such students. The difference in the former group is about 17 percentage points, whereas it is about minus two percentage points in the latter group. As Figure 7 shows, the difference becomes wider, as the percent of nonwhite students increases.

These moderation effects are also observed by turnover types. For instance, inexperienced principals (i.e., less than four years of experience) are more likely to transfer to different schools when they face the informal sanction, compared with those with the same experience level but not facing the sanction; the difference between the two groups is much wider among the most experienced principals, which is different from the pattern observed in the binary turnover analysis. Similarly, inexperienced principals are much more likely to change

105

their positions when they face informal sanction, whether it is the first time or not. The predicted probability of position changes among inexperience principals is nine percent when they face informal sanction for the first time, while it is two percent when they do not. The difference becomes small, or even turns into negative as the principals' experience level rises.

The formal NCLB sanctions do not appear to influence their turnover behaviors. I find no evidence that facing SIY 1 is correlated with principal turnover, or that the relationship is moderated by years of principal experience or other principal qualifications. Neither do I find evidence that SIY 2 or sanctions beyond SIY 2 are associated with principal turnover. However, because the number of principals in the sample facing these sanctions is relatively small, especially those for SIY 2 and beyond, these null findings may not suggest that the relationship is non-existent.

Finally, I examine whether the patterns in transfer and position changes are systematically different by sanction status. Although I find no evidence that that is the case, I find some notable patterns. First, principals tend to transfer from Title I schools to non-Title I schools, whether they face informal sanction or not. Along with this pattern, they tend to move to schools enrolling a smaller number of students who are eligible for the federal free/reduced lunch program. I also find that principals do not appear to consider any distance to the AYP threshold when deciding where to move. The distance variable changes very little before and after transfer. For position changes, principals tend to take central office positions, whether they face the informal sanction or not. Around 40 percent of principals who changed their positions took jobs at the central office. Their position's full time equivalency also tends to be reduced to less than 0.75. On the other hand, those facing informal sanction are more likely to become assistant principals or teachers when compared to those who are not facing it.

106

These results suggest that policymakers should provide professional support for inexperience principals who are working at low-performing schools, especially if they face some kind of sanctions. These principals may not possess skills and experience to turn around persistently low-performing schools. Support may include mentoring by experienced principals, assigning more professional staff, such as reading coaches and curriculum specialists, providing more funding to start new educational programs, and offering the principals some professional development targeted for school improvement, among others. As discussed earlier, frequent principal turnover negatively affects school operations and instructions, which lead to a decline in student achievement (Béteille et al., 2012; Hargreaves & Fink, 2006; Meyer et al., 2009; Miller, 2013). Longer stability could lead to more successful school reform. On the other hand, it may be good for low-performing schools that inexperienced principals leave their schools if the new ones come with stronger qualification. This does not appear to be the case in Missouri. Low-performing schools are not successful in hiring principals with stronger qualifications (Grissom & Mitani, 2016).

This study faces several limitations. First, as discussed in the theory section, it examines the relationship between NCLB sanctions and principal turnover through a reduced form approach, which does not allow for making a distinction of voluntary turnover from involuntary turnover. Although I descriptively provided some evidence that principal turnover might be largely initiated by principals themselves, the analysis does not examine principal performance measures. It is still possible that the district administrators systematically replaced low-performing principals with high-performing ones. Second, one of the key assumptions behind the propensity score matching is strongly ignorable treatment assignment assumption, or conditional independence assumption (CIA). This means that the assignment to the treatment (e.g., informal

sanction, SIY 1) is independent of the outcomes (i.e., principal turnover), conditional on the covariates. A threat to this assumption is unobservable confounders that are correlated with both the outcomes and the treatment. While the empirical models control for as many observable principal, school, and district characteristics as possible, my estimates can still be biased due to the unobservable factors, such as support from parents and local community organizations, school climate toward learning, and school funding to start educational programs. Use of the distance variable in the matching model would reduce bias from these factors, but would not be able to remove it completely.

Another limitation is that this study focuses on turnover at the end of the treatment year. It is possible that principals may have left their schools a couple of years after facing the sanction. To the extent that this is true, my estimates fail to capture the true effect of NCLB sanctions. Future research can examine the influence of NCLB sanctions dynamically using the same set of matched samples, but through a survival model which tracks the same principals over time.

My estimation models are also subject to anticipation effects (see Malani & Reif, 2010). If principals who are not facing any sanctions changed their turnover behaviors based on the anticipation that they would face sanctions near future, it may bias my parameter estimates downwardly, because principals in the control group also responded to the sanction, making turnover rates between the two groups more similar. Finally, small sample sizes affected statistical power and my empirical approaches for the second research question. More data would help estimate the association more precisely.

The literature would benefit from future work that replicates the study by using different data sets. Given that Missouri is a rural state, it is worth investigating whether the results hold in larger states with more urban areas. It would also benefit from studies that examine principal

turnover from both the demand and supply sides simultaneously and distinguish involuntary turnover from voluntary turnover. Future work might also include some studies that use a dynamic approach to investigate a cumulative influence of the sanctions.

References

Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina*. NBER Working Paper No. 20511. Retrieved from http://www.nber.org/papers/w20511. Cambridge, MA: National Bureau of Economic Research.

Akiba, M. & Reichardt, R. (2004). What predicts the mobility of elementary school leaders? An analysis of longitudinal data in Colorado. *Education Policy Analysis Archives*, *12*(18), 1-20.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*(2), 150-161.

Baker, B., & Cooper, B. S. (2005). Do principals with stronger academic backgrounds hire better teachers? Policy implications for high poverty schools. *Educational Administration Quarterly*, *41*(3), 449-479.

Baker, B., Punswick, E., & Belt, C. (2010). School leadership stability, principal moves, and departures: Evidence from Missouri. *Educational Administration Quarterly, 46*(4), 523–557.

Ballou, D., & Springer, M. G. (2011). *Has NCLB encouraged educational triage? Accountability and the distribution of achievement gains*. Unpublished manuscript. Nashville, TN: Vanderbilt University.

Bandeira de Mello, V. (2011), *Mapping state proficiency standards onto the NAEP scales: Variation and change in state standards for reading and mathematics, 2005–2009* (NCES 2011-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC: Government Printing Office.

Barry, S. (2015, January 9). Former Robert Hughes academy principal gets 1 year of probation for MCAS cheating scam at Springfield charter school. *MASSLIVE*. Retrieved on September 17, 2015, from http://www.masslive.com/news/index.ssf/2015/01/former_robert_m_hughes_academy.html.

Berman, M. (2015, April 1). Former educators found guilty in Atlanta school cheating scandal. *Washington Post*. Retrieved on September 17, 2015, from https://www.washingtonpost.com/news/post-nation/wp/2015/04/01/nearly-a-dozen-found-guilty-in-atlanta-school-cheating-trial/.

Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The effects of education accountability on teachers: Are policies too-stress provoking for their own good? *International Journal of Education Policy and Leadership, 4*(5). Retrieved on July 19, 2014, from http://www.ijepl.org.

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, *41*, 904-919.

Boser, U., Wilhelm, M., & Hanna, R. (2014). *The power of the Pygmalion effect: Teachers expectations strongly predict college completion*. Washington, DC: Center for American Progress.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals*. NBER Working paper No. 17803. Retrieved from http://www.nber.org/papers/w17803. Cambridge, MA: National Bureau of Economic Research.

Brewer, D. J. (1993). Principals and student outcomes: Evidence from U.S. high schools. *Economics of Education Review*, *12*(4), 281-292.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, *163*(12), 1149-1156.

Brooks, K., Adams, S. R., & Morita-Mullaney, T. (2010). Creating inclusive learning communities for ELL students: Transforming school principals' perspectives. *Theory into Practice*, *49*(2), 145-151.

Brown, P. F. (2006). Preparing principals for today's demands. *Phi Delta Kappan*, *87*(7), 525-526.

Center on Education Policy. (2006). *Ten big effects of the No Child Left Behind Act on public schools*. Washington, DC: Center on Education Policy. Retrieved from www.cep-dc.org.

Chakrabarti, R. (2014). Incentives and responses under No Child Left Behind: Credible threats and the role of competition. *Journal of Public Economics*, *110*, 124-146.

Clark, D., Martorell, P., & Rockoff, J. (2009). *School principals and school performance*. Working paper 38. Washington, DC: The Urban Institute.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, *31*(1), 92-109.

Cullen, J. B. & Mazzeo, M. J. (2008). *Implicit performance awards: An empirical analysis of the labor market for public school administrators*. Working paper.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418-446.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252-279.

Dhuey, E., & Smith, J. (2013). *How school principals influence student learning*. Unpublished manuscript.

Diamond, J. B., Randolph, A., & Spillane, J. P. (2004). Teachers' expectations and sense of responsibility for student learning: The importance of race, class, and organizational habitus. *Anthropology and Education Quarterly*, *35*(1), 75-98.

Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review*, *7*(3), 291-299.

Educational Testing Service. (2008). *Standards, accountability, and flexibility: Americans speak on No Child Left Behind reauthorization*. Princeton, NJ: Educational Testing Service.

Farley-Ripple, E. N., Solano, P. L., & McDuffie, M. J. (2012). Conceptual and methodological issues in research on school administrators career behavior. *Educational Researcher*, *41*(6), 220-229.

Fruehwirth, J. C., & Traczynski, J. (2013). *Spare the rod? The dynamic effects of failing accountability on schools*. Unpublished manuscript. Cambridge, UK: University of Cambridge.

Fuller, E., & Young, M. (2009). *Tenure and retention of newly hired principals in Texas.* Austin: University Council of Educational Administration, Department of Educational Administration, University of Texas at Austin.

Gabriel, T. (2010, June 10). Under pressure, teachers tamper with tests. *New York Times*. Retrieved on September 17, 2015, from http://www.nytimes.com/2010/06/11/education/11cheat.html?_r=0.

Gardiner, M. E., & Enomoto, E. K. (2006). Urban school principals and their role as multicultural leaders. *Urban Education*, *41*(6), 560-584.

Gardiner, M. E., Canfield-Davis, K., & Anderson, K. L. (2009). Urban school principals and the 'No Child Left Behind' Act. *The Urban Review*, *41*(2), 141-160.

Gates, S. M., Ringel, J. S., Santibañez, L., Guarino, C., Ghosh-Dastidar, B., & Brown, A. (2006). Mobility and turnover among school principals. *Economics of Education Review*, *25*(3), 289-302.

Grissom, J.A. & Andersen, S. (2012). Why superintendents turn over. *American Educational Research Journal*, *49*(6), 1146-1180.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433-444.

Grissom, J. A., & Mitani, H. (2016). *Principal sorting and the distribution of principal quality: Documenting inequities in state and national data*. Working paper. Nashville, TN: Vanderbilt University.

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers and their work environment. *Educational Evaluation and Policy Analysis, 36*(4), 417-436.

Gutman, L. M., & McLoyd, V. C. (2000). Parents' management of their children's education within the home, at school, and in the community: An examination of African-American families living in poverty. The *Urban Review*, *32*(1), 1-24.

Hallinger, P. & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*(2), 157-191.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*(3), 234-249.

Hargreaves, A., & Fink, D. (2004). The seven principles of sustainable leadership. *Educational Leadership*, *61*(7), 8-13.

Hargreaves, A., & Fink, D. (2006). *Sustainable leadership*. San Francisco, CA: Jossey Bass.

Hargreaves, A., Moore, S., Fink, D., Brayman, C., & White, R. (2003). *Succeeding leaders? A study of principal succession and sustainability*. Toronto, Ontario: Ontario Principal's Council.

Hemelt, S. W. (2011). Performance effects of failure to make Adequate Yearly Progress (AYP): Evidence from a regression discontinuity framework. *Economics of Education Review*, *30*(4), 702-723.

Jencks, C., & Phillips, M. (1998). *The black-white test score gap*. Washington, DC: The Brookings Institute.

Johnson, J. V., & Hall, E. M. (1988). Job strain, work place social support, and cardiovascular disease: A cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health*, *78*(10), 1336-1342.

Karasek, R. A. Jr. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, *24*(2), 285-308.

Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis*, *34*(2), 185-208.

Lee, J., & Bowen, N. K. (2006). Parent involvement, cultural capital, and the achievement gap among elementary school children. *American Educational Research Journal*, *43*(2), 193-218.

Li, D. (2012). *School accountability and principal mobility: How No Child Left Behind affects the allocation of school leaders*. Working paper. Cambridge, MA: Harvard University.

Loeb, S., Kalogridges, D., & Horng, E. L. (2010). Principal preferences and the even distribution of principals across schools. *Educational Evaluation and Policy Analysis*, *32*(2), 205-229.

Lyons, J. E., & Algozzine, B. (2006). Perceptions of the impact of accountability on the role of principals. *Education Policy Analysis Archives*, *14*(16), 1-16.

Malani, A., & Reif, J. (2010). *Accounting for anticipation effects: An application to medical malpractice tort reform.* NBER Working Paper No. 16593. Cambridge, MA: National Bureau of Economic Research.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.

McGhee, M. W., & Nelson, S. W. (2005). Sacrificing leaders, villainizing leadership: How educational accountability policies impair school leadership. The *Phi Delta Kappan*, *86*(5), 367-372.

Mertler, C. (2011). Teachers' Perceptions of the Influence of *No Child Left Behind* On Classroom Practices. *Current Issues in Education*, *13*(3). Retrieved on July 20, 2014, from http://cie.asu.edu/

Meyer, M. J., & Macmillan, R. B., Northfield, S. (2009). Principal Succession and Its Impact on Teacher Morale. *International Journal of Leadership in Education*, *12*(2), 171-185.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, *36*(1), 60-72.

Missouri National Education Association. (2012). *Tenure frequently asked questions.* Retrieved on April 4, 2016, from http://mnea.org/Missouri/LegalServices1/Tenure-Frequently-Asked-Questions-8.aspx.

Mitchell, T. R., & Lee, T. W. (2001). The unfolding model of voluntary turnover and job embeddedness: Foundations for a comprehensive theory of attachment. *Research in Organizational Behavior*, *23*, 189-246.

Missouri Department of Elementary and Secondary Education. (2010). *Understanding your adequate yearly progress (AYP) report 2010-11*. Jefferson City, MO: Missouri Department of Elementary and Secondary Education.

Reback, R., Rockoff, J., Schwartz, H. L., & Davidson, E. (2011). Barnard/Columbia No Child Left Behind Database, 2002-03 and 2003-04. http://www.gsb.columbia.edu/nclb.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy, 6*(3), 207-241.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169-188.

Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record*, *107*(9), 1999-2045.

Salazar, K. (2007). *Senator Salazar's No Child Left Behind survey results*. Retrieved on September 17, 2015, from https://web.archive.org/web/20070131231836/http://salazar.senate.gov/other/nclbresults.html#who.

Sanders, E. T. W. (1999). *Urban school leadership: Issues and strategies*. Larchmont, NY: Eye on Education.

Santoro, D. A. (2011). Good teaching in difficult times: Demoralization in the pursuit of good work. *American Journal of Education*, *118*(1), 1-23.

Schnall, P. L., Lansbergis, P. A., & Baker, D. (1994). Job strain and cardiovascular disease. *Annual Review of Public Health*, *15*, 381-411.

Shipps, D. (2012). Empowered or beleaguered? Principals' accountability under New York City's diverse provider regime. *Education Policy Analysis Archives*, *20*(1). Retrieved on June 4, 2016, from http://epaa.asu.edu/ojs/article/view/892.

Solano, P. L., McDuffie, M. J., & Farley-Ripple, E. N., & Bruton, J. (2010). *Principal retention in the state of Delaware.* Newark, DE: University of Delaware, Center for Community Research and Service.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1-21.

Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom realities and No Child Left Behind*. The Civil Rights Project. Cambridge, MA: Harvard University.

Taylor, L. L., Glander, M., & Fowler, W. J. (2007). *Documentation for the NCES Comparable Wage Index Data File, 2005* (EFSC 2007-397). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Texas Association of School Board. (2008). A comparison of the NCLB and Texas accountability systems. Retrieved on December 4, 2014, from https://www.tasb.org/Legislative/Issue-Based-Resources/documents/account_diff2008.aspx.

United States Department of Education. (2002). *No Child Left Behind: A desktop reference 2002*. Washington, DC: United States Department of Education.

Van der Doef, M., & Maes, S. (1999). The job demand-control (-support) model and psychological well-being: A review of 20 years of empirical research. *Work & Stress*, *13*(2), 87-114.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced Leadership: What 30 years of research tells us about the effect of leadership on student achievement* (Working paper). Aurora, CO: Mid-continent Research for Education and Learning.

White, B. R., & Agarwal, P. K. (2011). *The principal report: The State of School Leadership in Illinois*. Edwardsville, IL: Illinois Education Research Council, Southern Illinois University Edwardsville.

Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioural health of children and youth. *American Psychologist*, *67*(4), 272-284.

Supply Side (Principals)
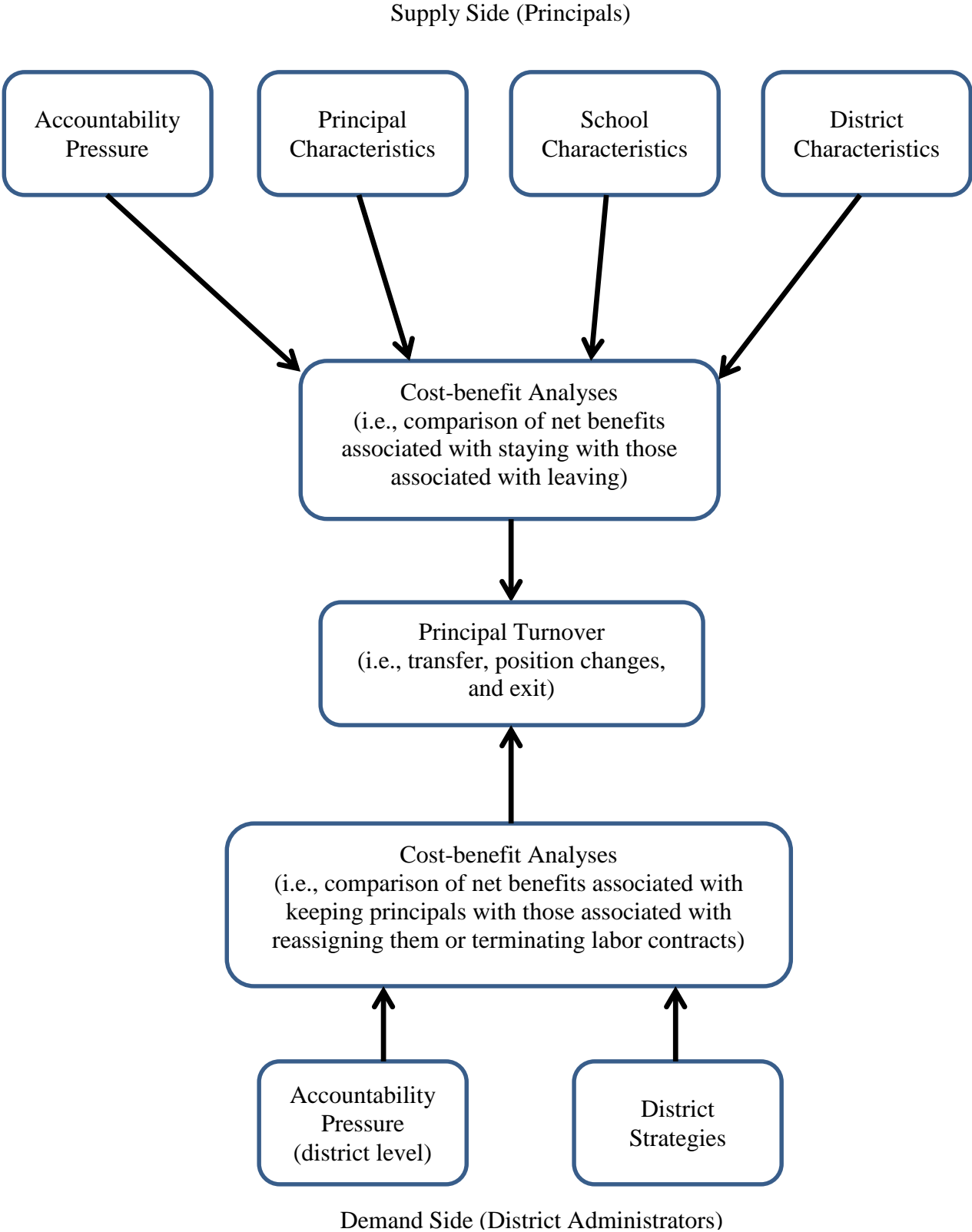
Figure 1: Conceptual framework

Demand Side (District Administrators)
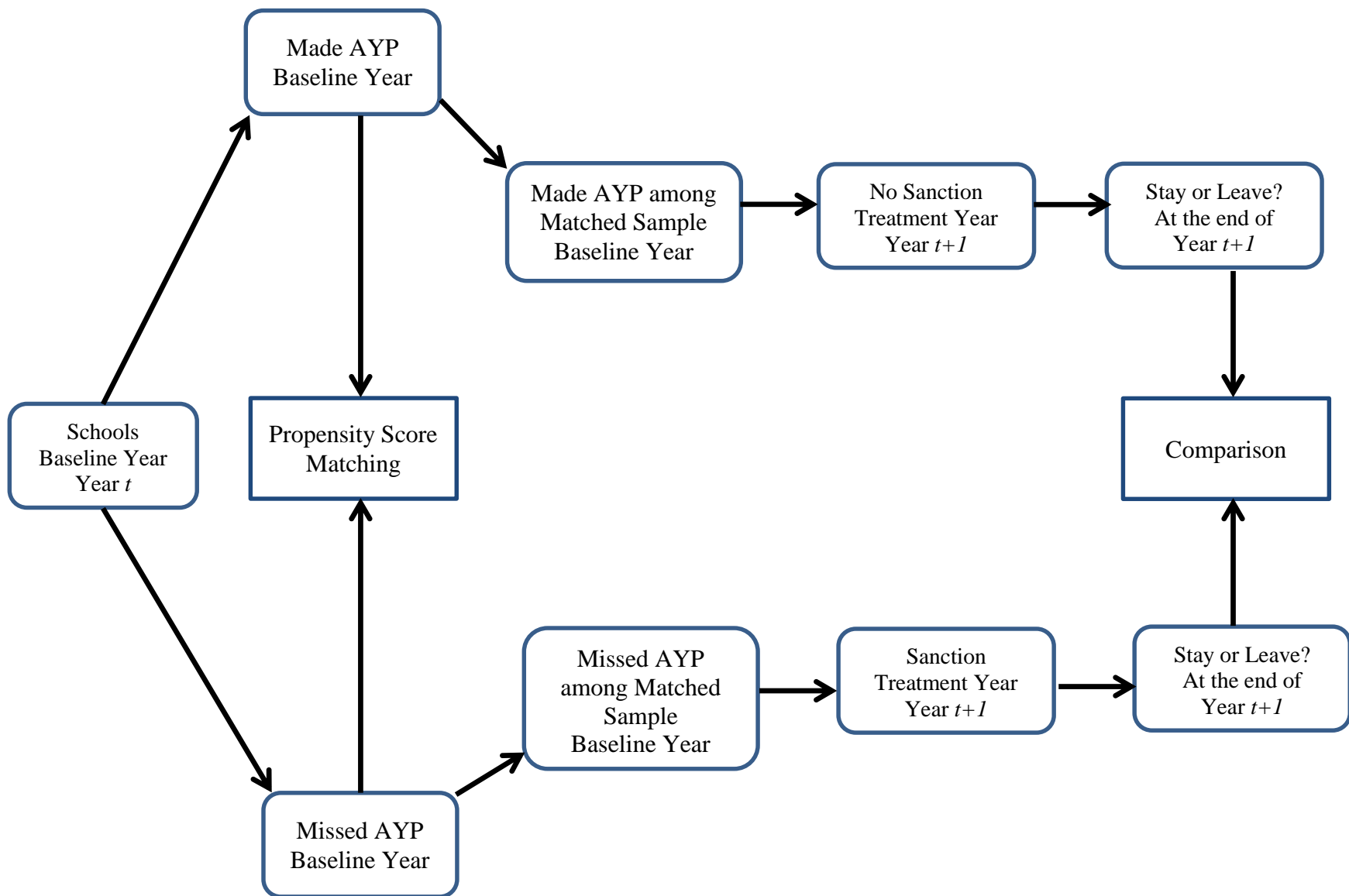
Figure 2: Implementation of the propensity score matching
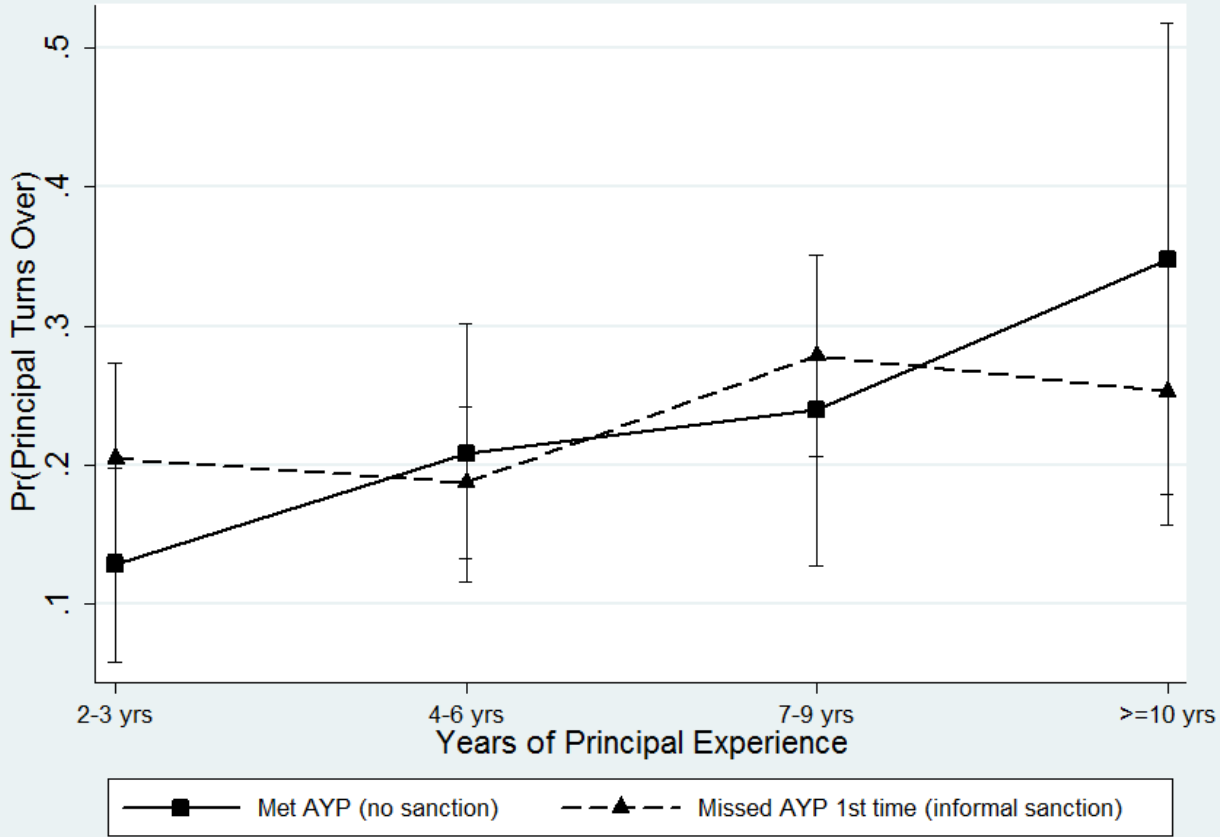
Figure 3: Predicted principal turnover by years of principal experience
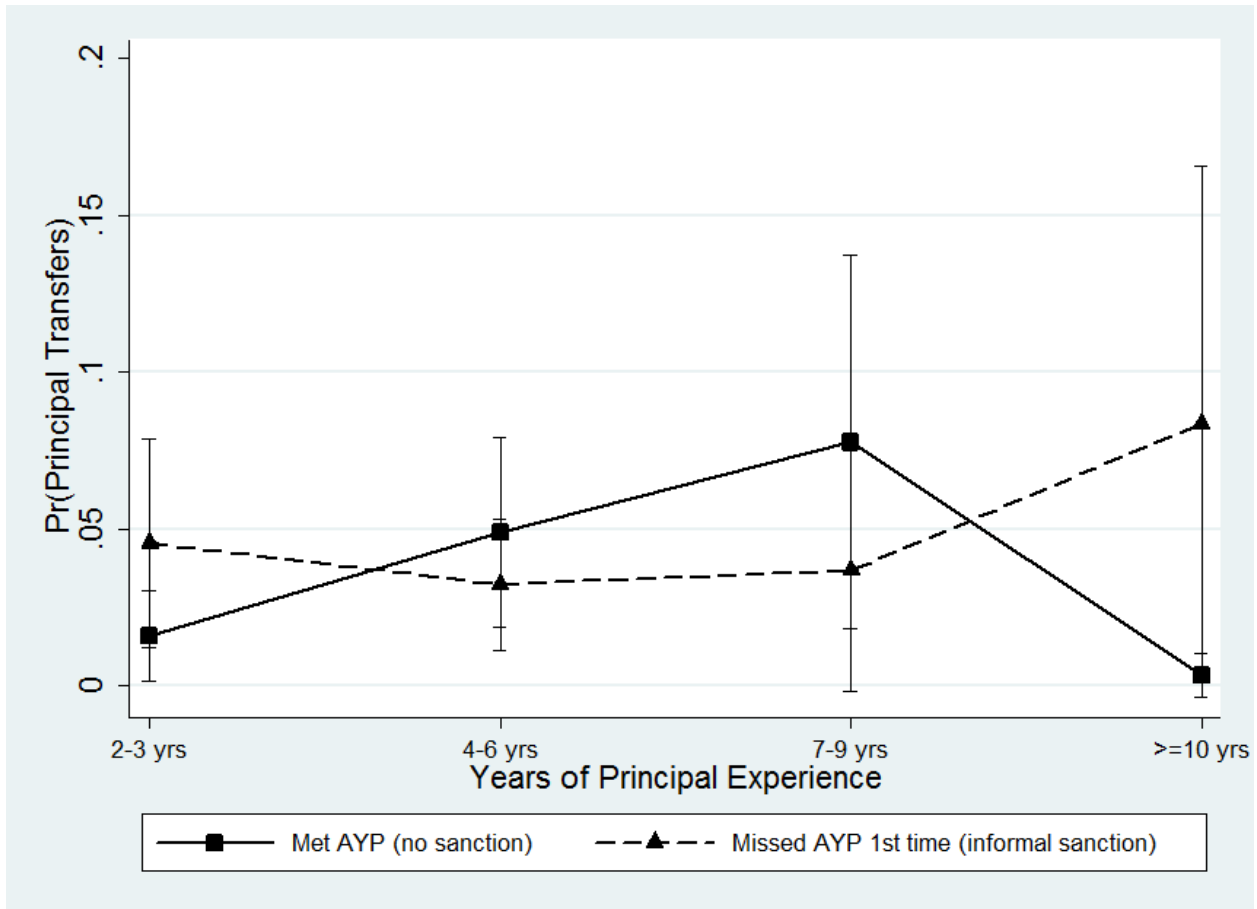
Figure 4: Predicted principal transfer by years of principal experience
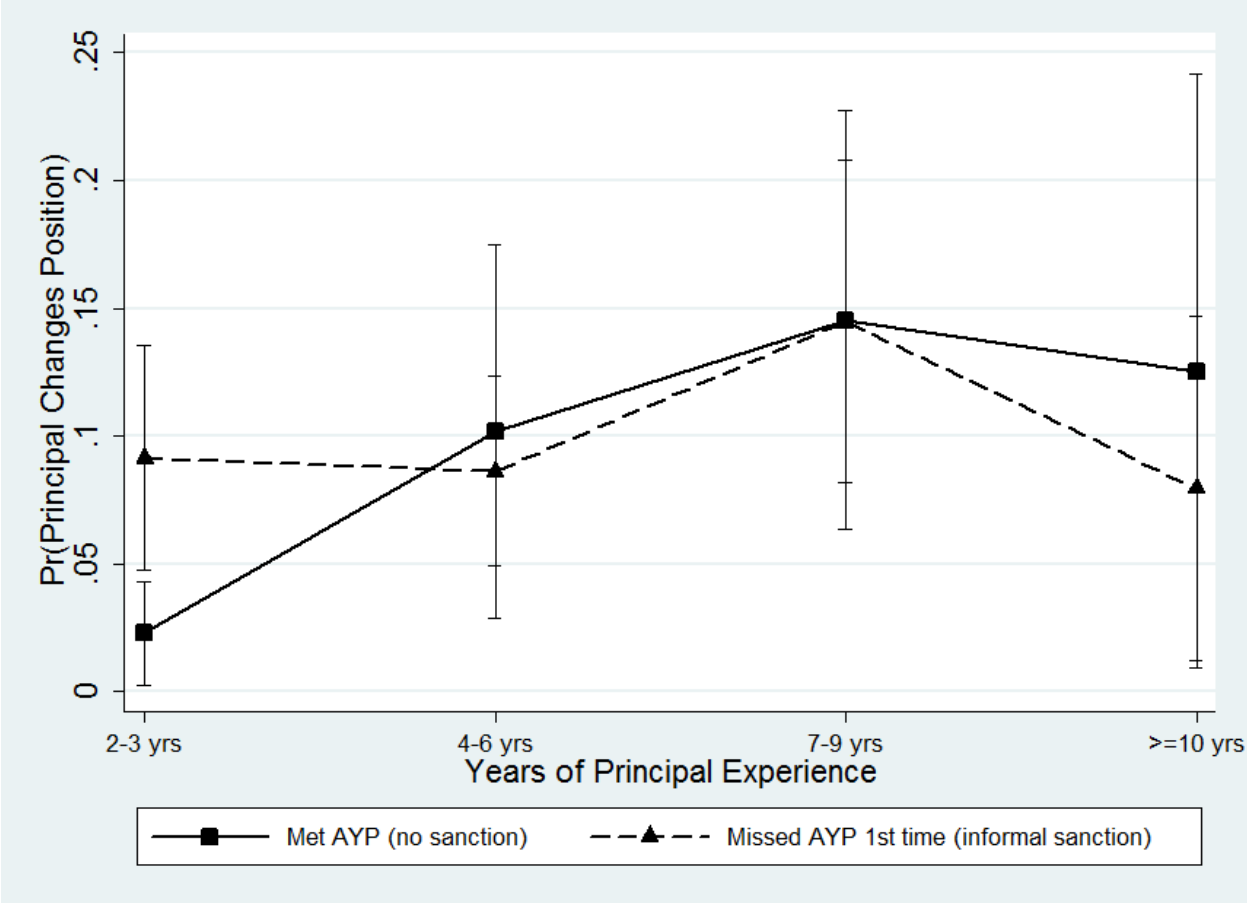
Figure 5: Predicted probability of position changes by years of principal experience
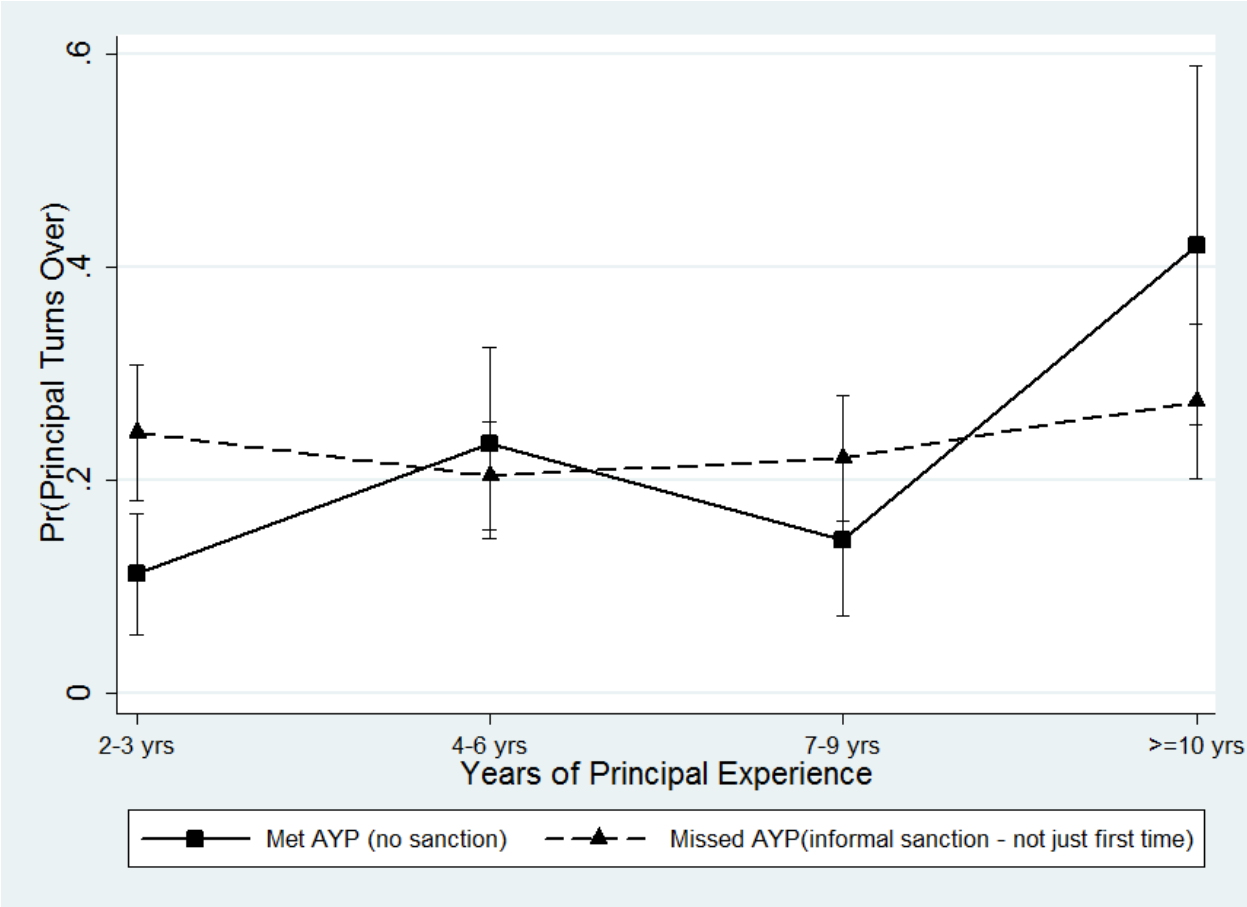
Figure 6: Predicted principal turnover by years of principal experience
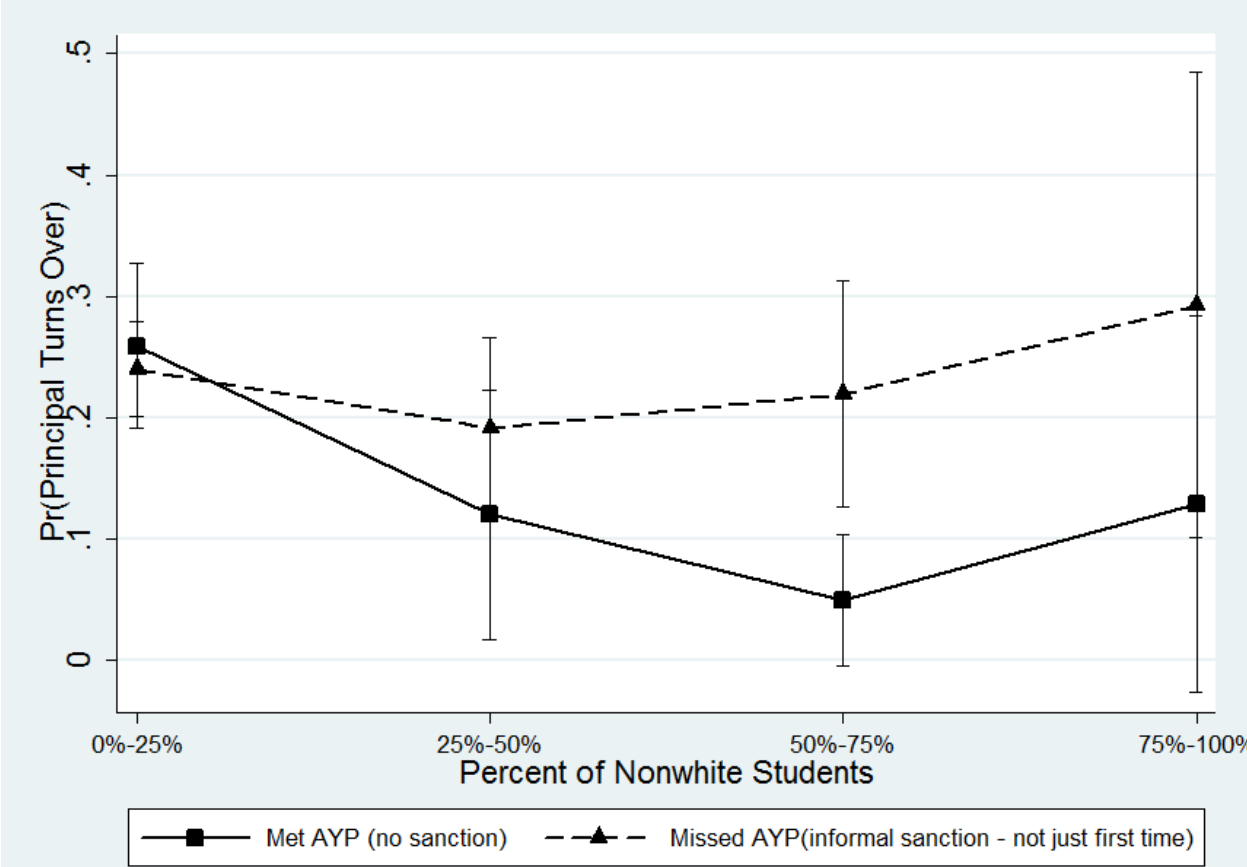
Figure 7: Predicted principal turnover by percent of nonwhite students

Table 1: NCLB sanction system

Number of _consecutive_ years that school missed AYP in same subject in at the end of year _t-1_	Sanction imposed in the _following_ school year (year _t_)
0	No sanction imposed.
1	No sanction imposed but the failure is publicly announced.
2	School Improvement Year 1 (SIY 1). District must offer Public School Choice (PSC) with transportation to all students to transfer to another public school or charter school within the same district which has not been identified for "in-need-of improvement. It also needs to provide technical assistance.
3	School Improvement Year 2 (SIY 2). District must offer supplemental educational services (SES) to students from low-income families, in addition to PSC and technical assistance.
4	School Improvement Year 3 (SIY 3). District must take at least one of the following corrective actions on failing schools: (1) replace school staff, (2) implement a new curriculum (with appropriate professional development), (3) decrease management authority at the school level, (4) appoint an outside expert to advise the school, (5) extend the school day or year, and (6) reorganize the school internally. District continues to offer PSC, SES, and technical assistance.
5	School Improvement Year 4 (SIY 4). District must initiate plans to fundamentally restructure failing schools. It continues to offer PSC, SES, and technical assistance.
6	School Improvement Year 5 (SIY 5). District must implement the school restructuring plan. It includes one or more of the following actions: (1) reopen the school as a public or charter school, (2) replace all or most of the staff who are responsible for the failure to make AYP, (3) enter into a contract with an entity, such as private management company, to operate the school, (4) turn the operation of the school over to the state department of education, and (5) other major restructuring. District continues to offer PSC, SES, and technical assistance.
7 and beyond	School remains in restructuring until it meets AYP for two consecutive years.

Notes: If a district does not have PSC options because all schools are identified for "in-need-of improvement," it offers SES instead. Data source: US Department of Education

Table 2: Annual measurable objectives (AYP targets) by subjects in Missouri

Year	Communication Arts (%)	Math (%)
2001-02	18.4	8.3
2002-03	19.4	9.3
2003-04	20.4	10.3
2004-05	26.6	17.5
2005-06	34.7	26.6
2006-07	42.9	35.8
2007-08	51.0	45.0
2008-09	59.2	54.1
2009-10	67.4	63.3
2010-11	75.5	72.5
2011-12	83.7	81.7
2012-13	91.8	90.8
2013-14	100	100
Participation rate	All schools	95
Attendance rate	Elementary/Middle schools	93
Graduation rate	High schools	85

Notes: These are school-level targets. The USDOE started requiring additional academic indicators from 2005-06.

Table 3: Descriptive statistics

Covariate	No sanctions baseline year			
	No sanction treatment year		Informal sanction treatment year	
	N	Mean	N	Mean
Principal characteristics				
Age	3244	46.46	1014	47.36***
Female	3244	0.61	1014	0.47***
Black	3244	0.04	1014	0.12***
Other nonwhite	3244	0.01	1014	0.00**
Attended selective undergraduate institutions	3222	0.18	1011	0.19
Education specialist or doctoral degree	3244	0.38	1014	0.38
Total years of principal experience	3244	6.21	1014	6.04
Years in current school as principal	3244	4.58	1014	4.60
Total years of experience in education	3244	18.24	1014	19.27***
Relative salary ratio	3221	1.01	1000	1.01
School characteristics				
Distance to making AYP (percent)	3236	28.86	1009	3.99***
Urban	3244	0.16	1014	0.22***
Suburban	3244	0.31	1014	0.29
Town	3244	0.09	1014	0.15***
Rural	3244	0.44	1014	0.35***
Elementary school	3244	0.83	1014	0.47***
Middle school	3244	0.08	1014	0.23***
High school	3244	0.08	1014	0.26***
Other grade configuration	3244	0.01	1014	0.04***
Title I school	3244	0.66	1014	0.54***
Percent nonwhite students	3244	13.46	1009	25.10***
Percent low-income students	3244	37.06	1009	44.74***
School enrollment size	3244	378	1009	505***

District characteristics

Number of schools	3244	14.95	1014	22.35***

Notes: Data include years from 2001-02 to 2007-08. This sample is limited to principals whose schools had not missed AYP in the baseline year, which varies from principal to principal. Principals who did not stay in the same school in either the baseline year or the treatment year are not included. The last baseline year is 2007-08 because the treatment year is 2008-09 in this case, and I need data on 2009-10, the last year of the data, to identify whether principals turned over at the end of 2008-09. Both Title I and non-Title I schools are included. The means of these variables are compared using a series of t tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Principal turnover rates by sanction status in the pre-matching sample of principals whose schools had never missed AYP until the baseline year

	No sanction in the baseline year	
	No sanction treatment year	Informal sanction treatment year
Turnover types		
Stayers	81.45	75.43***
Movers	4.64	4.50
Within districts	2.28	2.31
Across districts	2.36	2.19
Position changers	7.79	10.71**
Leavers	6.12	9.37***

Notes: Data include years from 2001-02 to 2007-08. This sample is limited to principals whose schools had not missed AYP in the baseline year, which varies from principal to principal. Principals who did not stay in the same schools either the baseline year or the treatment year are not included. The last baseline year is 2007-08 because the treatment year is 2008-09 in this case, and I need data on 2009-10, the last year of the data, to identify whether principals turned over at the end of 2008-09. Both Title I and non-Title I schools are included. The means are compared using a series of t tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Balance test results for the sample of principals whose schools had never missed AYP until the baseline year

Variable	No sanction in the baseline year		
	Facing no sanction treatment year	Facing informal sanction treatment year	Standardized bias (%)
Principal characteristics			
Age	45.8	45.6	3.2
Female	0.50	0.59	-18.6
Black	0.09	0.06	10.7
Other nonwhite	0.00	0.00	0.3
Attended selective undergraduate institutions	0.18	0.25	-19
Education specialist or doctoral degree	0.37	0.33	8.2
Total years of principal experience	5.24	5.46	-6.3
Years in current school as principal	3.74	4.14	-15.1
Total years of experience in education	17.67	17.49	2.3
Relative salary ratio	1.01	1.04	-21.1
School characteristics			
Distance to making AYP (percent)	5.44	7.07	-12.30
Urban	0.17	0.20	-7.70
Town	0.16	0.08	23.80
Rural	0.38	0.30	17.00
Middle school	0.23	0.21	5.20
High school	0.24	0.20	10.90
Other grade configuration	0.02	0.01	7.10
Title I school	0.55	0.50	10.90
Percent nonwhite students	19.93	21.28	-5.30
Percent low-income students	43.02	37.53	24.20
School enrollment size	460	512	-16.80
District characteristics			
Number of schools	17.21	19.11	-7.5

Notes: Data include years from 2001-02 to 2007-08. The matching is based on a group of principals whose schools had not missed AYP yet as of the baseline year. I also included interaction terms, and their standardized biases are less than 10 in absolute value.

Table 6: Association between first time failure (informal sanction) and principal turnover – results from logistic regression

	Logistic regression			Logistic regression with PSM		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Facing informal sanction	1.23*	1.04	1.21	1.07	1.13	1.03
	(1.87)	(0.22)	(1.57)	(0.34)	(0.34)	(0.15)
Sanction x Title I		1.24			0.91	
		(1.04)			(-0.23)	
Sanction x high school			1.10			1.20
			(0.35)			(0.35)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observation	3529	3319	3529	1512	1512	1512
Pseudo R-squared	0.03	0.03	0.03	0.09	0.09	0.09

Notes: Principals who did not stay in the same school in either the baseline year or the treatment year are not included. All models include characteristics of principals, schools, and districts. Odds ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Moderation effects by principal qualifications

Panel A: Principal Qualifications	Model 1	Model 2	Model 3	Model 4
Facing informal sanction	1.07	1.18	1.78	1.84
	(0.28)	(0.66)	(1.28)	(1.57)
Sanction x selective college	1.03			
	(0.05)			
Sanction x education specialist/doctorate		0.79		
		(-0.66)		
Sanction x total years of principal experience			0.92	
			(-1.37)	
sanction x total years of principal experience (4 to 6 years)				0.47
				(-1.42)
sanction x total years of principal experience (7 to 9 years)				0.68
				(-0.73)
sanction x total years of principal experience (10 years or longer)				0.33*
				(-1.88)
Year fixed effects	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes
Observation	1512	1512	1512	1512
Pseudo R-squared	0.09	0.09	0.10	0.10

Panel B: Student Demographics	Percent nonwhite		Percent FRL	
	Model 1	Model 2	Model 3	Model 4
Facing informal sanction	1.12	1.06	1.94	1.72
	(0.47)	(0.27)	(1.22)	(0.92)
Sanction x percent of nonwhite students/FRL students	1.00		0.99	
	(-0.35)		(-1.32)	
Sanction x percent of nonwhite students/FRL students (25%-50%)		2.26		0.62
		(1.28)		(-0.72)
Sanction x percent of nonwhite students/FRL students (50%-75%)		0.41		0.57

		(-1.12)		(-0.85)
Sanction x percent of nonwhite students/FRL students (75%-100%)		0.80		0.36
		(-0.32)		(-1.19)
Year fixed effects	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes
Observation	1512	1512	1512	1512
Pseudo R-squared	0.09	0.12	0.10	0.09

Notes: Principals who did not stay in the same school in either the baseline year or the treatment year are not included. FRL stands for the federal free and reduced lunch. All models include characteristics of principals, schools, and districts. Odds ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Association between informal sanction and principal turnover rates by type

	Model 1			Model 2			Model 3		
	Movers	Changers	Leavers	Movers	Changers	Leavers	Movers	Changers	Leavers
Facing informal sanction	1.23	1.23	1.00	0.85	1.73	0.86	1.30	1.34	0.82
	(0.52)	(0.74)	(0.00)	(-0.25)	(1.17)	(-0.26)	(0.60)	(1.06)	(-0.58)
Sanction x Title I				1.78	0.51	1.32			
				(0.79)	(-1.21)	(0.41)			
Sanction x high school							0.75	0.74	4.66
							(-0.37)	(-0.50)	(1.25)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1512			1512			1512		
Pseudo R-squared	0.18			0.18			0.18		

Notes: New principals in current schools are excluded from the analysis. All models include characteristics of principals, schools, and districts. A variable for other race was dropped because of the convergence problem. Relative risk ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Moderation effects on principal turnover types by years of principal experience

Panel A: College selectivity and highest degree	Model 1			Model 2		
	Movers	Changers	Leavers	Movers	Changers	Leavers
Facing informal sanction	1.23	1.30	0.92	1.46	1.68	0.96
	(0.48)	(0.91)	(-0.23)	(0.77)	(1.59)	(-0.10)
Sanction x college selectivity	1.01	0.77	1.40			
	(0.01)	(-0.40)	(0.48)			
Sanction x highest degree (ed. specialist or doctoral)				0.71	0.50	1.20
				(-0.53)	(-1.46)	(0.29)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1512			1512		
Pseudo R-squared	0.18			0.18		

Panel B: Years of principal experience	Model 3			Model 4		
	Movers	Changers	Leavers	Movers	Changers	Leavers
Facing informal sanction	1.09	2.96**	1.39	3.46*	4.69***	0.53
	(0.13)	(2.14)	(0.41)	(1.88)	(2.85)	(-1.08)
Sanction x total years of principal experience	1.02	0.86**	0.95			
	(0.24)	(-2.31)	(-0.49)			
Sanction x total years of principal experience (4 to 6 years)				0.18*	0.17**	3.22
				(-1.90)	(-2.35)	(1.36)
Sanction x total years of principal experience (7 to 9 years)				0.13**	0.21**	9.51**
				(-1.97)	(-2.34)	(2.56)
Sanction x total years of principal experience (10 years or longer)				8.33	0.13**	0.73
				(1.58)	(-2.27)	(-0.40)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1512			1512		
Pseudo R-squared	0.18			0.21		

Notes: New principals in current schools are excluded from the analysis. All models include characteristics of principals, schools, and districts. A variable for other race was dropped because of the convergence problem. Relative risk ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Principal turnover rates by turnover type and sanction status in the pre-matching sample

	No sanction in the baseline year		Informal sanction in the baseline year	
	No sanction next year	Informal sanction next year	No sanction next year	SIY 1 next year
Stayers	80.74	75.17***	77.97	77.23
By Type				
Movers	4.82	3.79	5.24	3.40*
Within-districts	2.29	1.42*	1.75	1.05
Across-districts	2.52	2.37	3.50	2.35
Position changers	8.73	11.28**	11.54	11.35
Leavers	5.71	9.76***	5.24	8.02**

Notes: Data include years from 2001-02 to 2007-08. The first sample includes all principals who did not face any sanction in the baseline year, whether they worked at Title I or non-Title I schools. The second sample is limited to principals working at Title I schools, who faced informal sanction in the baseline year. Principals who did not stay in the same school in either the baseline year or the treatment year are not included. The last baseline year is 2007-08 because the treatment year is 2008-09 in this case, and I need data on 2009-10, the last year of the data, to identify whether principals turned over at the end of 2008-09. The means are compared using a series of t tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Association between informal sanction and SIY 1 and principal turnover

Panel A: Informal sanction	Model 1	Model 2	Model 3
Facing informal sanction	1.12	0.84	1.06
	(0.59)	(-0.62)	(0.28)
Sanction x Title I		1.94**	
		(1.97)	
Sanction x high school			1.22
			(0.45)
Year fixed effects	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes
Observation	2074	2074	2074
Pseudo R-squared	0.10	0.11	0.10

Panel B: SIY 1	Model 4	Model 5
Facing SIY 1	0.94	0.90
	(-0.16)	(-0.26)
Sanction x high school		1.41
		(0.31)
Year fixed effects	Yes	Yes
Labor region fixed effects	Yes	Yes
Observation	384	384
Pseudo R-squared	0.32	0.32

Notes: New principals in current schools are excluded from the analysis. All models include characteristics of principals, schools, and districts. For Panel B, the sample includes Title I school principals only. Schools whose SIY 1 status is delayed are not included. Odds ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Moderation effects by years of principal experience and percent of nonwhite students– informal sanction and SIY 1

Panel A: Informal sanction

	Years of principal experience		Percent nonwhite students	
	Model 1	Model 2	Model 3	Model 4
Facing informal sanction	2.62**	2.75***	0.80	0.89
	(2.40)	(3.15)	(-0.95)	(-0.53)
Sanction x yrs prin exp / pct nonwhite	0.89**		1.02***	
	(-2.16)		(2.98)	
Sanction x yrs prin exp (4 to 6 years) / pct nonwhite (25%-50%)		0.30***		2.05
		(-2.70)		(1.24)
Sanction x yrs prin exp (7 to 9 years) / pct nonwhite (50%-75%)		0.64		6.89***
		(-0.87)		(3.34)
Sanction x yrs prin exp (10 years or longer) / pct nonwhite (75%-100%)		0.18***		3.51***
		(-3.39)		(3.28)
Year fixed effects	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes
Observation	2074	2074	2074	2074
Pseudo R-squared	0.11	0.13	0.11	0.11

Panel B: SIY 1

	Years of principal experience		Percent nonwhite students	
	Model 5	Model 6	Model 7	Model 8
Facing SIY 1	0.84	1.02	1.10	1.03
	(-0.23)	(0.04)	(0.21)	(0.06)
Sanction x total years of principal experience	1.02		1.00	
	(0.18)		(-0.37)	
Sanction x total years of principal experience (4 to 6 years)		1.73		2.20
		(0.54)		(0.73)
Sanction x total years of principal experience (7 to 9 years)		0.62		0.01***
		(-0.56)		(-2.78)
Sanction x total years of principal experience (10 years or longer)		0.77		1.08
		(-0.29)		(0.09)

Year fixed effects	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes
Observation	384	384	384	384
Pseudo R-squared	0.32	0.33	0.32	0.35

Notes: New principals in current schools are excluded from the analysis. All models include characteristics of principals, schools, and districts. For Panel B, the sample includes Title I school principals only. Schools whose SIY 1 status is delayed are not included. Odds ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Association between informal sanction and SIY 1 and principal turnover by type

Panel A: Informal sanction	Model 1			Model 2			Model 3		
	Movers	Changers	Leavers	Movers	Changers	Leavers	Movers	Changers	Leavers
Facing informal sanction	0.87	1.07	1.68*	0.60	1.23	0.98	1.13	1.16	1.22
	(-0.40)	(0.31)	(1.74)	(-0.98)	(0.67)	(-0.05)	(0.39)	(0.66)	(0.61)
Sanction x Title I				2.16	0.71	3.30**			
				(1.39)	(-0.84)	(2.11)			
Sanction x high school							0.39	0.80	6.08**
							(-1.44)	(-0.50)	(2.34)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	2074			2074			2074		
Pseudo R-squared	0.20			0.21			0.21		

Panel B: SIY 1	Model 4			Model 5		
	Movers	Changers	Leavers	Movers	Changers	Leavers
Facing SIY 1	2.99	0.77	1.54	2.35	0.72	1.49
	(1.41)	(-0.56)	(0.74)	(0.99)	(-0.68)	(0.58)
Sanction x high school				3.13	3.41	1.27
				(0.74)	(0.60)	(0.19)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observation	384			384		
Pseudo R-squared	0.48			0.48		

Notes: New principals in current schools are excluded from the analysis. All models include characteristics of principals, schools, and districts. For Panel B, the sample includes Title I school principals only. Schools whose SIY 1 status is delayed are not included. An indicator variable for South Central is dropped because few principals in the region were matched. Odds ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Principal turnover rates by type and sanction status from SIY 1

	Baseline year		Baseline year		Baseline year		Baseline year	
	SIY 1		SIY 2		SIY 3		SIY 4	
	Next year		Next year		Next year		Next year	
	SIY 1 delayed	SIY 2	SIY 2 delayed	SIY 3	SIY 3 delayed	SIY 4	SIY 4 delayed	SIY 5
Number of principals	34	98	16	53	1	26	0	3
Stayers	76.47	66.33	81.25	62.26	100	53.85	0	33.33
By Type								
Movers	8.82	9.18	12.50	7.55	0.00	11.54	0.00	0.00
Within-districts	2.94	5.1	6.25	7.55	0.00	7.69	0.00	0.00
Across-districts	5.88	4.08	6.25	0.00	0.00	3.85	0.00	0.00
Position changers	5.88	12.24	0.00	13.21***	0.00	19.23	0.00	66.67
Leavers	8.82	12.24	6.25	16.98	0.00	15.38	0.00	0.00

Notes: Only Title I schools are included. Schools facing any delayed sanction are excluded. Principals who did not stay in the same school in either the baseline year or the treatment year are not included. The last baseline year is 2007-08 because the treatment year is 2008-09 in this case, and I need data on 2009-10, the last year of the data, to identify whether principals turned over at the end of 2008-09. The means are compared using a series of t tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Falsification test

Panel A: First time informal sanction

	Model 1		Model 2	
	Turnover	Movers	Changers	Leavers
Facing informal sanction	1.24	1.03	1.46*	1.14
	(1.50)	(0.14)	(1.81)	(0.51)
Year fixed effects	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes
Observations	1979	1980		
Pseudo R-squared	0.03	0.09		

Panel B: Informal sanction

	Model 3		Model 4	
	Turnover	Movers	Changers	Leavers
Facing informal sanction	0.91	0.91	0.94	0.79
	(-0.73)	(-0.50)	(-0.28)	(-1.17)
Year fixed effects	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes
Observations	2269	2269		
Pseudo R-squared	0.02	0.09		

Panel C: SIY 1

	Model 5		Model 6	
	Turnover	Movers	Changers	Leavers
Facing informal sanction	1.28	1.22	1.65	0.52
	(0.93)	(0.48)	(1.28)	(-0.79)
Year fixed effects	Yes	Yes	Yes	Yes
Labor region fixed effects	Yes	Yes	Yes	Yes
Observations	571	584		
Pseudo R-squared	0.07	0.19		

Notes: The year of the data is from 1993-94 to 1999-2000. New principals in current schools are excluded from the analysis. All models include characteristics of principals, schools, and districts. Odds ratios are reported. z statistics are reported in parentheses. Standard errors are clustered at the district level but not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Changes in school characteristics before and after transfer by sanction status (no sanction and all informal sanction)

Panel A: All schools — No sanction in the baseline year

| | | No sanction in the treatment year | | | | | Informal sanction in the treatment year | | | | | |
	N	Sending schools	Receiving schools	Diff. (R-S)	p value	N	Sending schools	Receiving schools	Diff. (R-S)	P value	$D_{informal} - D_{nosanction}$	P value
Distance to making AYP	47	18.85	19.57	0.72	0.82	25	14.75	15.01	0.26	0.95	-0.46	0.93
Percent low-income students	55	42.31	37.02	-5.29	0.09	27	49.41	45.78	-3.62	0.29	1.67	0.72
Percent nonwhite students	57	15.75	17.49	1.74	0.52	34	17.63	20.33	2.70	0.28	0.96	0.77
School enrollment size	55	382	540	158	0.00	27	393	485	93	0.16	-66	0.34
Title I schools (percent)	57	0.58	0.46	-0.12	0.11	34	0.56	0.44	-0.12	0.29	0.01	0.97
Elementary/middle schools	57	0.81	0.82	0.02	0.66	34	0.76	0.76	0.00	1.00	-0.02	0.78
High schools	57	0.18	0.16	-0.02	0.57	34	0.21	0.24	0.03	0.66	0.05	0.52

Panel B: Title I schools — No sanction in the baseline year

| | | No sanction in the treatment year | | | | | Informal sanction in the treatment year | | | | | |
	N	Sending schools	Receiving schools	Diff. (R-S)	p value	N	Sending schools	Receiving schools	Diff. (R-S)	p value	$D_{informal} - D_{nosanction}$	P value
Distance to making AYP	26	20.17	21.08	0.91	0.84	13	13.90	17.56	3.66	0.50	2.75	0.69
Percent low-income students	29	48.16	37.29	-10.88	0.02	15	57.40	47.39	-10.00	0.05	0.87	0.89
Percent nonwhite students	29	22.13	22.54	0.41	0.91	20	25.84	28.31	2.47	0.54	2.06	0.71
School enrollment size	29	419	495	76	0.11	15	399	469	69	0.41	-6	0.94
Title I schools (percent)	29	1.00	0.62	0.38	0.00	20	0.95	0.50	0.45	0.00	-0.07	0.61
Elementary/middle schools	29	1.00	1.00	0.00	1.00	20	0.90	0.85	0.05	0.58	0.05	0.57
High schools	29	0.00	0.00	0.00	1.00	20	0.10	0.15	-0.05	0.58	-0.05	0.57

Notes: School characteristics for sending schools are based on the data in the treatment year; school characteristics for receiving schools are based on data for next year. Principals without valid data for both receiving and sending schools are excluded. Principals who did not stay in the same school in either the baseline year or the treatment year are not included. I used bivariate regressions for continuous variables to perform statistical tests. For the binary variables, I used paired t-tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Position principals took by sanction status (no sanction and all informal sanction)

| Position | No sanction in the baseline year | | |
	No sanction in the treatment year	Informal sanction in the treatment year	Total
Principal	18.52	20.59	19.52
Assistant principal	8.33	12.75	10.48
Teacher	12.04	14.71	13.33
Central office	42.59	39.22	40.95
Other school administrator	12.04	6.86	9.52
Supervisor	4.63	2.94	3.81
Other	1.85	2.94	2.38
Total	100	100	100

Chi-squared proportion test: *Chi-squared* = 3.6447, p = 0.725

Notes: If the full time equivalency changes from 0.75 or above to below 0.75, I treat it as a position change.

Table 18: Descriptive analysis of changes in principal qualifications before and after turnover

Panel A: First time informal sanction

	Not facing first time informal sanction			Facing first time informal sanction			
	Leaving principal	New principal	Difference	Leaving principal	New principal	Difference	*D informal - D nosanction*
College selectivity	0.20	0.19	-0.01	0.27	0.15	-0.13**	-0.12*
Highest degree attained	0.39	0.27	-0.12**	0.41	0.28	-0.13**	-0.01
Years of principal experience	6.66	2.87	-3.79***	6.28	3.27	-3.01***	0.78

Panel B: Informal sanction

	Not facing informal sanction			Facing informal sanction			
	Leaving principal	New principal	Difference	Leaving principal	New principal	Difference	*D informal - D nosanction*
College selectivity	0.18	0.17	-0.01	0.23	0.18	-0.05	-0.04
Highest degree attained	0.43	0.29	-0.14***	0.38	0.26	-0.12**	0.01
Years of principal experience	6.64	2.61	-4.02***	6.73	2.82	-3.90***	0.12

Panel C: SIY 1

	Not facing SIY 1			Facing SIY 1			
	Leaving principal	New principal	Difference	Leaving principal	New principal	Difference	*D SIY 1 - D nosanction*
College selectivity	0.20	0.20	0.00	0.15	0.12	-0.03	-0.03
Highest degree attained	0.36	0.28	-0.08	0.52	0.27	-0.24**	-0.16
Years of principal experience	6.44	2.68	-3.76***	6.97	2.55	-4.42***	-0.66

Notes: Leaving principals include those who transferred to different schools, changed their positions, and/or exited the system. Duplicates are dropped. A series of t-tests and bivariate regression models are used to examine whether the differences are statistically distinguishable from zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 1: Data sources for AYP-related information

Year	Assessment Data	AYP	School Improvement
2002	NLSLSASD	DESE AYP Grid	NA
2003	Columbia/Barnard	Columbia/Barnard	NA
2004	Columbia/Barnard	Columbia/Barnard	DESE website
2005	DESE AYP data file	DESE AYP data file	DESE website
2006	DESE AYP data file	DESE AYP data file	DESE website
2007	DESE AYP data file	DESE AYP data file	DESE website
2008	DESE AYP data file	DESE school improvement data file	DESE school improvement data file
2009	DESE AYP data file	DESE school improvement data file	DESE school improvement data file

Notes: NA stands for not applicable.

Appendix Table 2: Descriptive statistics of the pre-matching samples facing no sanction or informal sanction or SIY 1 in the baseline year

| | No sanction in the baseline year | | | | Informal sanction in the baseline year | | | |
| | No sanction treatment year | | Informal sanction treatment year | | No sanction treatment year | | SIY 1 treatment year | |
Variable	N	Mean	N	Mean	N	Mean	N	Mean
Principal characteristics								
Age	4577	46.4	3424	47.8***	323	47.7	408	48.7*
Female	4577	0.56	3424	0.42***	323	0.64	408	0.56**
Black	4577	0.05	3424	0.16***	323	0.16	408	0.28***
Other nonwhite	4577	0.01	3424	0.00	323	0.00	408	0.00
Attended selective undergraduate institutions	4553	0.17	3416	0.20**	322	0.17	408	0.15
Education specialist or doctoral degree	4577	0.38	3424	0.40*	323	0.40	408	0.38
Total years of principal experience	4577	6.29	3424	6.17	323	6.58	408	6.30
Years in current school as principal	4577	4.66	3424	4.77	323	4.94	408	4.81
Total years of experience in education	4577	18.35	3424	19.43***	323	19.12	408	19.52
Relative salary ratio	4541	1.01	3396	1.03***	321	1.00	407	1.00
School characteristics								
Distance to making AYP (percent)	4569	26.76	3415	3.78***	323	24.63	408	3.21***
Urban	4577	0.15	3424	0.24***	323	0.22	408	0.35***
Suburban	4577	0.29	3424	0.30	323	0.25	408	0.24
Town	4577	0.11	3424	0.17***	323	0.15	408	0.14
Rural	4577	0.44	3424	0.29***	323	0.38	408	0.27***
Elementary school	4577	0.72	3424	0.32***	323	0.74	408	0.62***
Middle school	4577	0.11	3424	0.30***	323	0.14	408	0.25***
High school	4577	0.16	3424	0.34***	323	0.11	408	0.13
Other grade configuration	4577	0.01	3424	0.04***	323	0.01	408	0.01
Title I school	4577	0.61	3423	0.42***				
Percent nonwhite students	4577	15.27	3403	30.54***	323	29.36	408	46.07***

Percent low-income students	4577	39.60	3403	45.97***	323	54.84	408	60.19***
School enrollment size	4577	392	3404	617***	323	389	408	462***

District characteristics

Number of schools	4577	14.84	3424	23.76***	323	19.93	408	34.03***

Notes: Data include years from 2001-02 to 2007-08. Principals who did not stay in the same school in either the baseline year or the treatment year are not included. The last baseline year is 2007-08 because the treatment year is 2008-09 in this case, and I need data on 2009-10, the last year of the data, to identify whether principals turned over at the end of 2008-09. Both Title I and non-Title I schools are included in the first sample; only Title I schools are included in the second sample. The means of these variables are compared using a series of t tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 3: Balance test results for the 2nd research question

Variable	No sanction in the baseline year			Informal sanction in the baseline year		
	No sanction treatment year	Informal sanction treatment year	Standardized bias (%)	No sanction treatment year	SIY1 treatment year	Standardized bias (%)
Principal characteristics						
Age	46.1	47.5	-16.0	47.0	47.9	-11.2
Female	0.47	0.55	-16.5	0.55	0.60	-10.6
Black	0.10	0.06	14.0	0.20	0.20	1.1
Other nonwhite	0.00	0.00	1.3	0.00	0.00	.
Attended selective undergraduate institutions	0.19	0.24	-13.5	0.17	0.26	-23.2
Education specialist or doctoral degree	0.38	0.45	-13.9	0.47	0.47	0
Total years of principal experience	5.42	6.07	-17.3	5.78	6.31	-14.2
Years in current school as principal	3.94	4.52	-21.3	4.13	4.01	4.3
Total years of experience in education	18.03	18.76	-9.3	18.21	19.71	-18.8
Relative salary ratio	1.02	1.04	-14.1	0.99	1.01	-18.9
School characteristics						
Distance to making AYP (percent)	5.89	7.40	-11.5	7.01	7.50	-3.8
Urban	0.19	0.17	5.8	0.25	0.22	6.5
Town	0.17	0.11	19.2	0.16	0.08	20.9
Rural	0.36	0.31	11.4	0.35	0.39	-9.8
Middle school	0.24	0.16	21.8	0.24	0.16	20.9
High school	0.27	0.29	-6.2	0.12	0.08	15.3
Other grade configuration	0.02	0.01	6.0	0.01	0.03	-21.1
Title I school	0.50	0.45	10.6	NA	NA	NA
Percent nonwhite students	22.00	22.61	-2.5	34.80	35.05	-0.7
Percent low-income students	43.92	38.89	22.5	56.65	55.13	6.6
School enrollment size	501	571	-22.7	443	397	21.6
District characteristics						

Number of schools	18.16	17.10	4.4	23.16	20.30	8.3
Others						
Interaction terms/different forms of variables	NA	NA	NA	NA	NA	<|25|

Notes: Data include years from 2002 to 2008. Principals who did not stay in the same schools in either the baseline year or the treatment year are not included. The last baseline year is 2007-08 because the treatment year is 2008-09 in this case, and I need data on 2009-10, the last year of the data, to identify whether principals turned over at the end of 2008-09. Both Title I and non-Title I schools are included in the first sample; only Title I schools are included in the second sample. NA means (1) that the mean calculations are not applicable, (2) that interaction terms or different forms of variables are not included in the model, or (3) that the variable is not included in the model.

CHAPTER 4

PRINCIPALS' WORKING CONDITIONS, JOB STRESS, AND TURNOVER BEHAVIORS
UNDER NCLB ACCOUNTABILITY PRESSURE

Introduction

The federal No Child Left Behind Act was a performance-based accountability policy

that required all public schools to make Adequate Yearly Progress (AYP), defined by each state

education agency in reading and math for all measurable student subgroups every year. Schools

that failed to make AYP faced public scrutiny and criticism, as AYP results were publicly

announced at the end of each school year. Continuous failure led to sanctions including staff

replacement, state takeover, and school closure. The key assumption behind this federal policy

was that NCLB sanction threats and actual sanctions would provide incentives for educators and

school administrators to change their behaviors to avoid public criticism and sanctions.[47] It was

assumed that these changes would influence teachers' instructional practices and eventually raise

student performance.

Prior studies have recently investigated this assumption by examining the effects of

NCLB on students and teachers (e.g., Dee & Jacob, 2011; Grissom, Nicholson-Crotty, &

Harrington, 2014; Reback, Rockoff, & Schwartz, 2014). These studies generally find that NCLB

improved student performance, especially in mathematics, and that it lowered teachers'

perceptions of teacher cooperation and job security and shifted more time to specialist teachers in

high stakes subjects; yet, they find little evidence that NCLB affected teacher job satisfaction or

commitment.

---

[47] Henceforth, the term NCLB sanctions will be used to mean both NCLB sanctions and sanction threats.

Despite the number of NCLB studies on students and teachers, little research has systematically examined to what extent NCLB or NCLB sanctions have influenced school principals. As school leadership literature shows, school principals play a critical role in improving school and student outcomes (e.g., Branch, Hanushek, & Rivkin, 2012; Brewer, 1993; Grissom, Kalogrides, & Loeb, 2015; Grissom, Loeb, & Master, 2013; Hallinger & Heck, 1998; Waters, Marzano, & McNulty, 2003), with school leadership "second only to classroom instruction among school-related factors that affect student learning in school" (Wallace Foundation, 2013, p. 5). They work with teachers, parents, and local community organizations to improve classroom instruction and create supportive learning environments (Leithwood, Louis, Anderson, & Wahlstrom, 2004; Wallace Foundation, 2013; Waters et al., 2003). Understanding how principals responded to NCLB sanctions would provide useful information for policymakers to design accountability systems appropriately, whether at the federal, state, or local level.

Extant NCLB studies on school principals are largely qualitative studies, survey reports, and opinion essays. For example, survey reports show that a majority of principals viewed NCLB unfavorably and perceived it as an unfair system (Educational Testing Service, 2008; Lyons & Algozzine, 2006; Salazar, 2007). Anecdotal evidence suggests that NCLB made principalship even more challenging and that many principals could not meet the growing requirements by the law, causing a lot of job stress (Brown, 2006). As evidenced by high-stakes state accountability systems (McGhee & Nelson, 2005), NCLB might have created a culture of fear of test scores and led principals to be in an isolated position. These studies and reports suggest that NCLB changed school principals' working conditions and increased their job stress, which may have eventually led to turnover (McGhee & Nelson, 2005). However, its influence has yet to be systematically explored. Given the importance of school leadership (e.g., Branch et

152

al., 2012; Grissom et al., 2015; Hallinger & Heck, 1998; Waters et al.,2003), there is a demand for research that empirically examines the relationship between NCLB and school principals.

This study provides new evidence on this unexplored area using a nationally representative sample of principals from the Schools and Staffing Survey (SASS) 2007-08 and detailed school-level AYP data in 2006-07 systematically collected from 45 state education agencies. From these AYP data, I construct a variable to measure a distance to the AYP threshold and use it as a key matching variable in a propensity score matching model to identify comparable schools between those facing NCLB sanctions and those not facing them. As described in the next section, the NCLB sanction system was progressive and cumulative. While it would be important to examine whether severe sanctions such as school restructuring changed principals' behaviors, it would be equally important to investigate whether principals' behaviors changed once their schools entered the sanction system. It could be at this point that principals started feeling accountability pressure. On the other hand, long-time exposure to the NCLB sanctions may have made principals insensitive to them. For these reasons, this study focuses on principals whose schools just entered the sanction system in 2007-08. This includes principals facing the informal sanction (i.e., public criticism and scrutiny) and the first year sanction or School Improvement Year 1 (SIY 1). I combine these sanctions because the SASS data or the school-level AYP data does not identify which schools faced the informal sanction in 2007-08.

More formally, I first investigate (1-a) whether or not these NCLB sanctions are associated with principals' working conditions and job stress. Principals' responses could be different by their qualifications such as years of principal experience. Experienced principals might have had better skills to cope with increasing accountability pressure than inexperienced principals. Principals' responses could also vary by school level, Title I school status, and/or

153

student demographic characteristics. For instance, principals serving a large number of nonwhite students might have faced different challenges than their colleagues at schools with a large number of students from affluent families. So I examine (1-b) whether the association is moderated by principal qualifications and/or school characteristics. For principal characteristics, I focus on years of principal experience and the highest degree (i.e., education specialist degree or doctoral degree). For school characteristics, I examine Title I status, school level, the percent of students eligible for the federal free/reduced lunch (FRL) program, and the percent of nonwhite students.

Next, I investigate (2-a) whether the NCLB sanctions are correlated with higher principal turnover rates. Some principals might transfer to different schools after facing the sanctions; other principals could choose to leave the public education system or change their positions. To test these possibilities, I examine (2-b) whether the association is different by turnover types. Finally, I investigate (2-c) whether the association is moderated by principal qualifications and/or school characteristics.

Results from this study may provide useful implications for policymakers to enhance the effectiveness of the current accountability systems and incentive policies. If the study suggests that principals in general do not respond to a performance- or sanction-based incentive system, policymakers may have to consider changing the incentive structure, such as using rewards instead of sanctions and differentiating performance targets, taking school contexts into consideration. They may also need to consider using different performance measures, as principals' jobs have multiple dimensions. In addition, because many of the current education policies and programs use some form of incentives, knowing how principals respond to the performance- and sanction-based incentive in general could help policymakers identify who

154

needs what kind of job support. For instance, if principals tend to have more job stress when facing a sanction-based incentive, policymakers may offer them mentoring programs and/or assign additional school support staff to their schools. Similarly, if the study suggests that principals tend to turn over when facing a sanction-based incentive, policymakers may consider embedding a retention incentive program in an overall sanction-based incentive policy or program such that principals facing such sanction stay longer.

This paper proceeds as follows. The next section describes the NCLB sanction system. After that, I review the literature on the impact of NCLB on school principals and discuss how NCLB accountability pressure might influence principals' working conditions, job stress, and turnover behaviors. Then I describe data and methods used, followed by the result section. I conclude with the implications and limitations of this study.

NCLB Sanction System

General Description

Under the NCLB sanction system, all public schools were required to meet Adequate Yearly Progress (AYP) in reading and math, and the 95-percent test participation rate requirement in grades 3 to 8 and once at the high school level in all student subgroups every year.[48] Elementary and middle schools also had to meet the attendance rate requirement and high schools needed to meet the graduation rate requirement.[49] AYP was usually measured by the percent of students performing at or above the proficient level in state assessments in math and

---

[48] Most of the policy information discussed in this section is based on the U.S. Department of Education's (USDOE) desktop reference for NCLB (USDOE, 2002) unless indicated otherwise.

[49] The US Department of Education added a requirement for additional academic indicators in 2005-06. While many states chose to use attendance rates and graduate rates, other states used other academic indicators such as writing assessment scores and improvement in test scores.

reading.[50] State education agencies set an annual measurable objective (AMO) in each subject at the school level or the grade level, and all schools had to meet AMOs in all subjects in all subgroups to make AYP. The USDOE allowed state education agencies to use the confidence interval, safe harbor, student growth models, and other methods to increase statistical reliability of their AYP determinations.

The NCLB sanction system used AYP results to decide whether to place schools in the NCLB sanctions. Although not the formal sanction, the first stage sanction was public scrutiny and criticism. Schools faced this sanction when they failed to make AYP for the first time or for the first time after their sanctions were lifted. This was not an actual sanction, but generated pressure on schools and principals because AYP results were publicly announced. The second stage sanction was school choice. Schools receiving the federal Title I funding faced a threat of losing their students through school choice options when they failed to make AYP in the same subject for two consecutive years. At the same time, they received technical assistance from their districts. When they continued to miss AYP in the same subject, their districts had to offer supplemental educational resources to their students from low-income families in addition to school choice options and technical assistance. Sanctions became severer as schools continued to fail to make AYP. In the next stage, failing schools were required to take corrective actions, which included staff replacement, implementation of a new curriculum, and school reorganization. After this stage, schools were required to make a school restructuring plan and if they continued to fail, they had to implement it. Table 1 summarizes the sanction system. Schools facing the sanctions needed to make AYP in the same subject that caused sanctions for two consecutive years to exit them. Making AYP for one year only delayed their sanction status.

---

[50] Thirteen states use some form of proficiency index instead of the percent of proficiency achieved.

Variation in AYP Determination Rules across States

The USDOE provided state education agencies with great discretion over AYP determination rules. A prior study reports how differences in the rules were associated with schools' failure rates and suggests that small differences in the rules may cause notable differences in outcomes (Davidson, Reback, Rockoff, & Schwartz, 2015). I collected information on AYP determination rules from all state education agencies by reviewing the Consolidated State Application Accountability Workbooks and states' own AYP documents, and communicating with state education officers. Table 2 reports some of the determination rules across states. It shows that use of the confidence interval (CI) differed across states. The USDOE allowed states to use the CI to increase the reliability of AYP determinations. In 2006-07, 47 states used the CI. Among them, 22 states used 99% CI, whereas 17 states used 95% CI. Three states applied other confidence levels. Variation also exists in the formula used to calculate the CI (not reported), whether it was applied to AMOs or the actual performance, and whether it was applied to all student subgroups or the all-student group only (not reported).

In contrast, all states were allowed to use the safe harbor (SH). The SH was applied to schools that reduced the percent of students performing below the proficiency level by 10 percentage from the previous year. These schools could make AYP even if they missed it in the first place. States varied in terms of whether the CI was applied to the SH, and if so, the confidence level used, and whether the SH was applied to student subgroups only or all subgroups, including the all-student group (not reported). In 2006-07, 22 states applied the CI to the SH. All of these states but one used the 75 percent CI. The formula for the CI used in the SH also varied from state to state (not reported).

Furthermore, states were different in whether they used uniform averaging procedures and whether student growth models were used. Uniform averaging is a method that averages the current year's performance with previous years' to avoid the influence of yearly fluctuations. Twenty-eight states used some form of uniform averaging procedures. Nine states were approved by the USDOE to use student growth models. Although not reported, states also differed in whether the USDOE allowed states to use adjustments in the percent of proficiency achieved among students with disabilities.

Due to these and other not-reported wide variations in AYP determination rules, it is possible that two similar schools in terms of distance to the AYP threshold and principal and school characteristics, but located in different states, faced different AYP determinations. For example, one school might make AYP because their state used a 99 percent confidence interval, whereas the other school might not, because their state used the more restrictive 95 percent confidence interval. Similarly, one school might make AYP because their state was allowed to use the adjustment in performance for the disability subgroup, whereas the other school might not because the adjustment was not allowed. As described in detail in the method section, I make use of these variations in answering my research questions.

NCLB and School Principals

A key assumption behind the NCLB sanction system was that NCLB sanctions gave school principals incentives to raise student performance and avoid sanctions. Theoretically, principals were expected to change their leadership behaviors to influence classroom instructions and student learning. For example, facing the sanction pressure, principals might design academic curricula that focus on high-stakes subjects and allocate more time and resources to

them (e.g., Spillane & Kenney, 2012). They might spend more time on observing classroom teachers and advising them. They could also engage more teachers and other school staff in decision-making processes in key areas important for school improvement such as making school missions and vision, establishing academic standards and curriculum, hiring new teachers, and making school budget plans (e.g., Louis, Leithwood, Wahlstrom, & Anderson, 2010; Knapp, Kopland, Honig, Plecki, & Portin, 2010; Portin, Schneider, DeArmond, & Gundlach, 2003; Spillane, 2005; Spillane, Halverson, & Diamond, 2001; Wallace Foundation, 2013). Moreover, they might spend more time analyzing student-level data on test scores and other relevant information to evaluate teacher performance and decide future interventions (Knapp & Feldman, 2012; Koyama, 2014; Spillane, Diamond, Burch, Hallett, Jita, & Zoltners, 2002).

If these or other practices that principals engage in to avoid sanctions are quite different from their daily practices, principals may face extra work or have to work longer hours. They might also reallocate their time within the school day to new tasks or priorities they think will raise student achievement.

NCLB-induced changes in work demands may increase principals' job stress. The relationship between work demand and job stress has been widely examined in occupational and personnel psychology. Studies in these fields find that greater (psychological) job demands negatively affect one's well-being, increase their job stress level, and can even hurt physical well-being, especially when the person does not have work control or social support from his or her colleagues (e.g., Johnson & Hall, 1988; Karasek, 1979; Schnall, Landsbergis, & Baker, 1994; Van der Doef & Maes, 1999). Furthermore, when work demands outpace the time available to a person, his or her job stress level increases (Shuler, 1979).

Increased job stress can lead to principal turnover via occupational burnout. Burnout is a three-dimensional construct comprising exhaustion, cynicism, and inefficacy, among which exhaustion is the main facet of burnout and reflects its stress dimension (Maslach, Schaufeli, & Leiter, 2001). So when job stress increases, a person's feeling of exhaustion can also increase. Yet, exhaustion alone does not lead him or her to burn out because it does not capture all aspects of the relationship that the person has with his or her work. Exhaustion is accompanied by cynicism, which means that the person keeps distance emotionally and cognitively from his or her work, and inefficacy, which means his or her sense of ineffectiveness. Research finds that when a person burns out, he or she is more likely to have turnover intentions and eventually leave his or her employer (Leiter & Maslach, 2009).

Consistent with the idea that NCLB-induced job stress can lead to principal turnover, recent studies find that principals are more likely to turn over when they face NCLB sanctions. For example, Li (2012) finds that principals with high value-added tend to transfer to different schools with lower probabilities of failing to make AYP and that NCLB shifted the distribution of principal quality. Ahn and Vigdor (2014) find that the probability of principal exit from the public system increases by 6-18 percentage points when they face school restructuring, the highest level of the NCLB sanctions. Moreover, White and Agarwal (2011) report that schools that failed to make AYP tend to experience higher levels of principal turnover.

All of these studies suggest that NCLB changed principals' working conditions and increased their job stress level, which induced principals to leave their schools. However, principals' responses to NCLB sanctions might not be the same across all types of principals serving different types of schools. For example, it would be reasonable to think that experienced principals have a set of more effective leadership skills to cope with accountability pressure and

make effective leadership decisions than inexperienced, novice principals, who may feel more job stress and struggle with it, resulting in poorer school and student performance in state assessments (Dhuey & Smith, 2013; Shipps, 2012). In fact, prior work on principal effectiveness finds that principal experience is positively associated with student and school performance (Baker, Punswick, & Belt, 2010; Clark, Martorell, & Rockoff, 2009; Coelli & Green, 2012; Eberts & Stone, 1988; Fairley-Ripple, Solano, & McDuffie, 2012; Solano, McDuffie, Farley-Ripple, & Bruton, 2010). Similarly, principals with an advanced degree (i.e., education specialist degree and doctoral degree) may have different leadership skills to respond to accountability pressure, as these principals received additional, different kinds of leadership training in addition to what typical principal preparation programs provide.

Principals' responses could also be different according to the characteristics of schools that they serve. For example, by design, Title I school principals might feel greater pressure from NCLB sanctions than non-Title I school principals, because only Title I schools were subject to the sanctions. Similarly, elementary and middle school principals might feel more pressure than high school principals because more grades were tested at these school levels under the NCLB accountability system. Moreover, principals serving a large number of students from disadvantaged backgrounds might face more difficult educational challenges than their colleges serving students from more privileged families. For example, principals serving a culturally diverse student population would have to deal with language and multicultural issues (e.g., Brooks, Adams, & Morita-Mullaney, 2010; Gardiner, Canfield-Davis, & Anderson, 2009; Rogoff, 2003), work with parents with limited English proficiency to engage them in children's education (Gardiner et al., 2009), and support students in disruptive home environments (Gardiner & Enomoto, 2005). Similarly, principals serving a large number of low-income

161

students might face such challenges as low parental involvement (e.g., Gutman & McLoyd, 2000; Lee & Bowen, 2006), low academic expectations among teachers (e.g., Boser, Wilhelm, & Hanna, 2014; Diamond, Randolph, & Spillane, 2004), and mental and physical health problems (e.g., Yoshikawa, Aber, & Beardslee, 2012).

All of these potential differences in principal leadership skills by principal qualifications and challenges by school characteristics suggest that principals could feel different accountability pressure by these characteristics, resulting in different responses to the pressure. It is, therefore, worth investigating whether the relationship between NCLB sanctions and principals' behaviors are moderated by these characteristics. However, as discussed earlier, most of the extant NCLB studies on school principals examine principals' behaviors among small samples of school principals, provide theoretical arguments about how NCLB changed the school principalship, and/or pay little attention to potential moderation effects by principal and school characteristics. While these studies provide insights to understand principals' responses to NCLB sanctions, there remains the question of whether these findings can be generalizable to a broader population of principals and whether moderation effects exit. Little research has systematically investigated these questions. This study fills this hole in the literature and contributes to it by providing new empirical evidence on these areas.

Data

My analysis relies on school principals' responses to the Schools and Staffing Survey (SASS) in 2007-08 and the Principal Follow-up Survey (PFS) in 2008-09 administered by the National Center for Education Statistics (NCES).[51] The SASS is a survey of a nationally

---

[51] Although the most recent SASS and PFS were administered in 2011-12 and 2012-13, this paper does not use these data for the following reasons. First, most of the states set AMOs above 80 percent, which were too high for many

representative sample of teachers, principals, schools, and districts that asked a wide range of

questions related to such topics as teacher and principal characteristics, working conditions,

instructional practices, and basic student demographics. I merge the SASS data with the PFS data,

which followed up with schools in 2008-09 to ask about principals. I further merge these data

with the NCES school-level Common Core of Data for 2006-07 to supplement school

characteristics data.[52] In addition, I systematically collected detailed school-level assessment and

AYP data disaggregated by student subgroups and subjects in 2006-07 from 45 states to

construct a measure of distance to the AYP threshold. For this data collection, I first searched

school-level assessment data files in state education agencies' websites. When I could not find

the data files, I used school report card websites that the state education agencies maintain and

downloaded each school's report card. When the states' report card websites did not include

report card data in 2006-07, I contacted state officials and made formal or informal data requests.

When the data requests were rejected, I utilized school-level assessment data maintained by the

National Longitudinal School-Level State Assessment Score Database (NLSLSASD) created by

the American Institutes for Research (AIR). The NLSLSASD maintains the 2006-07 assessment

data for most of the states. The main purpose of their data collection was to collect school-level

assessment data for 4th and 8th grades and was not directly related to analyze AYP determinations

(Blankenship, personal communication, May 12, 2015). In addition, there were some

uncertainties about the definition of assessment variables that they created. For these reasons, I

---

schools. Second, many schools missed AYP in 2010-11 (55% in the SASS 2011-12 data). Third, NCLB was under the reauthorization process during the time period, which generated a sense of uncertainty among principals about whether the US Department of Education continues to use the same sanction system. All of these could have weakened incentives among principals, which would prevent my statistical models from identifying the relationship between NCLB sanctions and principal behaviors. On the other hand, AMOs were still set relatively low (mostly around 50 percent) in most of the states in 2007-08, only 24 percent of schools missed AYP in 2006-07, and it had been only six years since the implementation of NCLB. Thus, the SASS 2007-08 and PFS 2008-09 data are more appropriate to answer the research questions in this paper.

[52] As explained later, the 2006-07 school year is a base year in my analysis and use base year characteristic variables to match schools.

did not turn to their database until the last step. From these steps, I could collect assessment data

from 46 states. However, because the unit of Oklahoma's AMOs was completely different from

that of the other states, I decided to drop the state from the analysis. I could not obtain usable

assessment data from Arkansas, Main, Nebraska, or Vermont. Along with this data collection, I

collected schools' NCLB sanction status in 2006-07 from state education agencies or the 2005-

06 Consolidated State Performance Reports Part I posted on the U.S. Department of Education

website. I use the data on the sanction status in 2006-07 to identify schools facing no formal

sanctions in that school year.

Dependent Variables

From the SASS and PFS data, I operationalize working conditions through measuring the

work demand and principals' job stress. For the work demanded, I use survey items that asked

the number of hours per week that principals spent on school-related activities during the school

day, before and after school, and over the weekends, and the number of hours per week that

principals spent on interacting with students. For principals' job stress, I use five survey items

(e.g., "The stress and disappointments involved in serving as principal of this school aren't really

worth it," "If I could get a higher paying job, I'd leave education as soon as possible," "I think

about staying home from school because I'm just too tired to go."), each of which is on a 1-4

Likert scale, with 1 being strongly agree and 4 being strongly disagree. I reverse-coded these

responses and performed a factor analysis, which revealed a single factor for job stress

(Cronbach's $a = 0.74$),[53] and standardized the factor score. Positive values mean that principals

had greater job stress; negative values mean that they had less job stress. Appendix Table 1

provides a full list of survey items and factor loadings.

---

[53] The one-factor solution employs varimax rotation.

I create three turnover variables from the PFS 2008-09. The first variable is a binary variable taking a value of one if a principal did not return to his or her school in 2008-09 and a value of zero otherwise. The second variable is a categorical variable that takes a value of one if a principal returned to his or her school as a principal in 2008-09, a value of two if the principal transferred to a different school as a principal, and a value of three otherwise.[54] I call the last category an exit for simplicity. The third variable differentiates within-district transfer from across-district transfer, yielding a total of four turnover categories (i.e., stayer, within-district transfer, across-district transfer, and other).

Methods

To investigate my research questions, I first restrict the sample to principals facing no formal sanction and principals facing the informal sanction in 2006-07. Then, I compare principals facing the informal sanction or School Improvement Year 1 (SIY 1) with those not facing them in 2007-08 using their schools' AYP results in 2006-07 as an indicator for their sanction status in 2007-08. That is, if a principal did not face any sanction in 2006-07 and his or her school missed AYP at the end of that school year, the principal faced the informal sanction in 2007-08. Similarly, if a principal faced the informal sanction in 2006-07 and his or her school missed AYP at the end of the school year, the principal faced SIY 1 in 2007-08. As explained in the introduction section, I combine the informal sanction and SIY 1 because the SASS data or the school-level AYP data does not identify which schools faced the informal sanction in 2007-08.

---

[54] The third category includes principals who changed their positions within the same schools or transferred to different schools but did not work as principals. It also includes principals who moved to district administrative offices, retired, deceased, or left the K-12 public education systems.

A problem with the simple two-group comparison is that it will result in biased estimates, because schools that missed AYP could be quite different from schools that made AYP in both observable and unobservable ways. For example, there are schools that missed AYP in 2006-07 by a large margin, and these schools might be persistently low-performing schools. There are also schools that made AYP by a large margin, and they might have performed well above AMO targets every year. Comparing principals working at these extreme schools will not yield meaningful results. On the other hand, there would be statistically similar schools across states in both principal and school characteristics as well as the "degree" to which they made or missed AYP; yet, due to differences in AYP determination rules employed by state education agencies, some of them made AYP and other missed AYP. I use this variation to identify the relationship between NCLB sanctions and principal behaviors.

The key to identifying these two groups of schools is to measure the "degree" to which schools made or missed AYP. I operationalize this construct by measuring the distance from the actual performance to the AYP threshold or the AMOs set by state education agencies. Recent studies that investigate the impact of NCLB constructed a minimum distance variable to the AYP threshold across all student subgroups and subjects to estimate the impact of NCLB (Ahn & Vigdor, 2014; Chakrabarti, 2014; Fruehwirth, & Traczynski, 2013; Hemelt, 2011). Based on their distance variable with some modifications, I construct a distance variable in the following way. First, I calculate the distance from the actual performance to the AMOs in each subject (i.e., math and reading) for each student subgroup meeting the minimum cell size requirement in 2006-07. Then, I take the minimum value across the subjects and subgroups within schools and use it as the minimum distance to the AYP threshold. More formally, I define the distance

variable as follows:

$$P_{sr}^{min} = min\{P_{jksr} - AMO_{kr}\}, \qquad (1)$$

where P_{sr}^{min} is the minimum value for school s in state r across subjects and subgroups in the

distance from the actual performance P_{jksr} in student subgroup j in subject k at school s in state r

to the annual measurable objectives AMO_{kr} in subject k in state r. If this value is greater than zero,

it means that all student subgroups performed above the AMOs in both subjects. Yet, this does

not mean that the school made AYP because the distance variable does not incorporate the other

AYP requirements such as participation rates, attendance rates, and graduation rates. Schools

still miss AYP if they do not meet these requirements.[55]

This distance variable plays a key role in identifying comparable schools. Yet, it alone is

not sufficient to identify statistically similar schools. Schools with similar distance to the AYP

threshold could be still different in other dimensions. To improve the quality of matching, I use a

propensity score matching method. In this method, I model the probability that a school missed

AYP in 2006-07 as a function of the distance, principal characteristics, and school characteristics

in 2006-07. More formally, I estimate the following logistic regression model:

$$\Pr(missing\ AYP)_{isr07} = \frac{e^f}{1+e^f}, \qquad (2)$$

where

$$f = \beta_0 + P_{isr07}\beta_1 + S_{sr07}\beta_2 + \varepsilon_{isr07}.$$

The probability that a school s in which a principal i works in state r in 2006-07 misses

AYP is a function of principal characteristics P_{isr07} (i.e., age, race, gender, education specialist or

doctoral degree, years of experience as a principal, and tenure in the current school as a

principal), school characteristics S_{sr07} (i.e., distance to making AYP, fraction of nonwhite

---

[55] This distance variable does not incorporate the confidence interval, the safe harbor, or other performance adjustment methods because I rely on variation in AYP determination rules across states to identify the relationship.

students, fraction of students eligible for the federal free/reduced lunch program, school enrollment size, school level, locality, and Title I status), and a random error term ε_{isr07}.[56] I created principal characteristic variables in 2006-07 based on the 2007-08 SASS data. Although principal characteristics may not be directly related to the AYP status, I include them in the matching model because they are correlated with the outcome variables in the second stage model described below. These variables help increase the precision of the estimates in the second stage model (Brookhart, Schneeweiss, Rothman, Glynn, Avorn, & Stürmer, 2006; Stuart, 2010).

From this model, I calculate the predicted probabilities or propensity scores, and perform the nearest neighbor matching with replacement based on the scores.[57,58] I set a caliper width as a quarter (0.25) of the standard deviation of the logit of the propensity scores to ensure the quality of matching (Austin, 2011; Rosenbaum & Rubin, 1985). After the matching, I perform a balance test to ensure that all covariates are balanced between the two groups. For assessing matching quality, I utilize the standardized bias, which is the mean difference as a percentage of the average standard deviation (Rosenbaum & Rubin, 1985), and use 25 percent or below as a criterion to determine whether a balance is achieved (Harder, Stuart, & Anthony, 2010; Rubin, 2001; Stuart, 2010). If the standardized bias for a covariate is greater than 25 percent, I add different forms of the covariate and/or interact it with other covariates.[59]

---

[56] The logistic regression model does not include survey weights because the purpose of this method is to balance covariates in the sample, not the population (DuGoff, Schuler, & Stuart, 2014; Zanutto, 2006). Ridgeway, Kovalchik, Griffin, and Kabeto (2015) argue that survey weights should be incorporated in the propensity score model under certain circumstances, which do not apply to my data.

[57] I performed other types of matching and found that the one-to-one nearest neighbor matching yields the best quality matching in terms of the covariate balance and the number of schools in common support.

[58] The propensity score matching method generates matching weights, which were equivalent to the total number of times a school in the control group was matched with schools in the treatment group. I multiply these weights by the SASS final principal weights to create a new weight variable used in the regression analysis (DuGoff et al., 2014; Ridgeway et al., 2015; Wang, 2015).

[59] In the actual matching, I did not need to include different forms of covariations or interaction terms.

Based on the matched schools, I model the outcome variable in 2007-08 as a function of sanction status in 2007-08 (or equivalently AYP result in 2006-07), principal characteristics, and school characteristics in 2006-07 to adjust for remaining differences in the covariates between the two groups.[60] More formally, I estimate the following regression model:

$$Y_{isr08} = \beta_0 + Sanction_{sr08}\beta_1 + P_{isr07}\beta_2 + S_{sr07}\beta_3 + \varepsilon_{isr08}. \tag{3}$$

Working conditions (or job stress) Y_{isr08} of a principal i at school s in state r in 2007-08 is a function of sanction status $Sanction_{sr08}$, principal characteristics P_{isr07}, school characteristics S_{sr07}, and a random error term ε_{isr08}.[61] I use the same principal and school characteristics variables as the ones used in Equation 2, except that the distance variable is excluded in this equation to avoid statistical overcontrol.

When the dependent variable is a binary turnover variable (i.e., stay or leave) or a categorical turnover variable (i.e., stay, move, or leave), I add a variable about the number of schools in a district to Equations 2 and 3 because it is related to the availability of jobs to principals, which would affect principal turnover rates (Farley-Ripple et al., 2012). I model the probability of principal turnover as a function of sanction status, the same principal and school covariates in Equation 3, and the district covariate. More formally, I estimate the following logistic regression model:

$$\Pr(Turnover)_{isdr08} = \frac{e^f}{1+e^f}, \tag{4}$$

where

$$f = \beta_0 + Sanction_{sdr08}\beta_1 + P_{isdr07}\beta_2 + S_{sdr07}\beta_3 + D_{dr07}\beta_4 + \varepsilon_{isdr08}.$$

---

[60] Because the 2006-07 school year was the base year, I use the covariates from that school year.
[61] The model does not include district fixed effects because close to 100 percent of matches are made across school districts. In addition, the model does not include state fixed effects because my method exploits variation in AYP determination rules across states.

The probability that a principal i working at school s in district d in state r turns over at the end of the 2007-08 school year is a function of the sanction status $Sanction_{sdr08}$, principal characteristics P_{isdr07}, school characteristics S_{sdr07}, district characteristics D_{dr07}, and a random error term ε_{isdr08}. If the coefficient on the sanction status is greater than 1, it means that the sanction is associated with higher turnover rates. When I use a categorical turnover variable, I estimate the following multinomial logistic regression model with the same set of covariates:

$$\Pr(turnover\ type = j)_{isdr08} = \frac{e^{f_j}}{1+\sum_{k=2}^{3} e^{f_k}}, \text{ for } j > 1 \tag{5}$$

where

$$f_j = \beta_{j0} + Sanction_{sdr08}\beta_{j1} + P_{isdr07}\beta_{j2} + S_{sdr07}\beta_{j3} + D_{dr07}\beta_{j4} + \varepsilon_{isdr08}.$$

Although turnover is a two-sided event between principals and school districts, as Equations 4 and 5 indicate, I use reduced form models to estimate the relationship between the sanction and turnover behaviors. These models do not distinguish voluntary turnover from involuntary turnover. Involuntary turnover typically results from district administrators' human resource decisions.

My analysis does not include principals who did not stay in the same school in 2006-07 or 2007-08 because the propensity score matching is based on 2006-07 and uses principal characteristics in 2006-07. I cluster standard errors at the district level in all models.

Results

Descriptive Analysis

I first descriptively examine characteristics of schools and principals in the base year of the sample (i.e., 2006-07). As explained earlier, this sample includes only principals who did not face any formal sanction in 2006-07. The top part of Table 3 reports that the average age of

principals is 49, and 52 percent are female. About 85 percent are white; black and Hispanic principals account for 14 percent of the sample. The average years of principal experience is about eight years and the length of tenure is just over four years. Forty percent of principals hold either an education specialist or doctoral degree.

The middle part of Table 3 reports characteristics of the schools in which these principals worked. The average percent of nonwhite students is 34 percent and the average percent of students eligible for the federal free/reduced lunch program is 40 percent. Sixty-three percent of the schools in the sample are elementary schools, 20 percent are middle schools, and 15 percent are high schools. The average school enrollment size is 587 students. One-fifth of the schools are located in urban areas and one-third of them are located in suburban areas. Town and rural schools account for 46 percent of the schools. Close to two-thirds of schools are Title I schools. The bottom part reports that the average number of schools in a district is about 37.[62]

The last two columns of the table compare these characteristics between schools that made AYP and schools that missed AYP in 2006-07. Principal characteristics tend to be similar between the two groups with some exceptions in female, black, and tenure. In contrast, school characteristics are quite different between the groups. For example, the average percent of nonwhite students among schools making AYP is 32 percent, whereas it is 41 percent among failing schools. Similarly, the average percent of students eligible for the federal free/reduced lunch program among the former group is 39 percent but 45 percent among the latter. The proportion of schools by school level also differs between the two groups. Close to 70 percent of

---

[62] I examined whether this sample is statistically different from the sample that includes all states by a series of *t* tests. I found that the sample is very similar to the whole sample in most of the principal characteristics except female and American Indian. However, the sample is statistically different from the whole sample in many of the school characteristics. Appendix Table 2 reports results.

the former group are elementary schools, but only 38 percent among the latter group. The difference in the number of schools per district between the two groups is about 20, and schools that missed AYP tend to be located in districts operating a larger number of schools. These differences, especially in school characteristics, highlight the importance of the use of propensity score matching.

Next, I descriptively compare principals' working conditions and turnover behaviors in 2007-08 by AYP status in 2006-07 by a series of two sample t-tests. Table 4 shows the results. The left side of the table displays the means of the outcome variables for all schools and the right side shows the means by AYP status. The top part of the table shows that, on average, principals spend close to 60 hours on school-related activities per week. This number includes hours spent before and after school and on weekends. Among these hours, they spend about 21 hours on interacting with students. The outcome variable about principals' job stress is standardized. The mean is -0.04 standard deviations. By AYP status, although the total number of hours spent on school-related activities is very similar between the two groups, the total number of hours spent on student interactions is significantly greater among principals at failing schools than their counterparts. Notably, principals at failing schools have greater job stress than their counterparts by 0.14 standard deviations.

The bottom part of the table reports principals' turnover behaviors. On average, 21 percent of principals do not return to their schools next year as principals. Seven percent transfer to new schools as principals, and 13 percent exit the system or change their positions. Among transferred principals, about 57 percent (or four percent as a whole) move to new schools within districts; 43 percent make across-district moves. By AYP status, principals at failing schools are more likely to leave their schools by five percentage points after facing the NCLB sanctions than

172

their counterparts, although the difference is not statistically significant at the conventional level. Similarly, they are more likely to move to new schools as principals.

Post-matching Analysis

Now, I turn to the post-matching analysis. From this sub-section, each analysis is based on a sample of matched schools using the propensity score matching method described in the method section.

NCLB Sanction and Principals' Working Conditions (Research Question 1)

In this section, I examine whether the NCLB sanction (i.e., the informal sanction and SIY 1 combined) is associated with changes in principals' working conditions (RQ 1-a) and whether the association is moderated by principal qualifications and/or school characteristics (RQ 1-b). For RQ 1-b, I examine the following principal and school characteristics: years of principal experience, the highest degree (i.e., education specialist degree or doctoral degree), Title I status, school level, the percent of FRL students, and the percent of nonwhite students. The left side of Appendix Tables 3 and 4 report results from a balance test for these research questions. Note that Appendix Table 4 excludes principals at schools serving grades that cross school levels (i.e., combination schools). Both tables show that covariates are balanced between the treatment and control groups.

Table 5 reports regression estimates from Equation 3. Although all models include principal and school characteristics, I omit those coefficients on them for brevity. The job stress variable is standardized, and larger values mean greater stress. The table shows that, although the

173

coefficient is positive, facing the sanction is not statistically significantly associated with an increase in the total hours of work or hours spent on interactions with students.

One notable pattern is that facing the sanction statistically significantly increases principals' job stress level. It is associated with an increase of 0.24 standard deviations in the job stress level. The magnitude of this coefficient is large, suggesting that principals felt greater job stress under the NCLB sanction. The job stress level is -0.14 standard deviations for the average principal not facing the sanction, whereas it was 0.10 standard deviations for the average principal facing the sanction.

Next, I examine whether these patterns are moderated by principal qualifications. For this analysis, I add to Equation 3 interaction term(s) between the sanction status and each of the following principal characteristics separately: education specialist degree or doctoral degree, years of principal experience, and a set of three indicator variables for years of principal experience (i.e., 4 to 6 years, 7 to 9 years, and 10 years or longer) to test for non-linearity. Table 6 displays results. I find little evidence that principal qualifications moderate the relationship, with a few exceptions. For instance, holding an education specialist degree or a doctoral degree negatively moderates the relationship for the total hours spent on student interactions, whereas it positively moderates the relationship for job stress. Principals with such an advanced degree are much more likely to feel job stress when they face the sanction than their colleagues not facing it. Among principals without such a degree, the difference in the job stress by sanction status is statistically indistinguishable from zero. Yet, overall, little moderation effect is observed.

Next, I investigate moderation effects by school characteristics. I use Title I school status, school level (i.e., high school), the percent of FRL students, and the percent of nonwhite students. Table 7 reports results for the first two characteristics; Table 8 shows results for the last two

174

characteristics. While Table 7 shows no evidence that the relationship is moderated by Title I school status or school level, Table 8 reports some evidence, though partly suggestive, that principals serving a large number of nonwhite students and low-income students (75% or more) tend to feel greater job stress (Model 6 in Panel A and Model 12 in Panel B). To ease the interpretation of these results, I plot the predicted job stress (standardized) by sanction status over the four categories of the percent of nonwhite students and the percent of FRL students in Figures 1 and 2, respectively.

Figure 1 shows that the difference in the job stress level between principals facing the sanctions and those not facing them becomes wider at high percentages of nonwhite students. The difference in the first category is 0.29 standard deviations and statistically significant. It becomes negative in the second category, but increases to 0.27 standard deviations in the third category. The difference is 0.62 standard deviations in the last category and the widest across the categories. Although not significant and thus suggestive, the difference in the differences is 0.33 standard deviations (p=0.107) and suggests that principals serving the largest number of nonwhite students tend to feel greater job stress than those serving the smallest number.

I observe similar trends for the percent of FRL students. Figure 2 shows that the difference in the job stress level between principals facing the NCLB sanctions and those not facing them gradually increases. The difference is 0.13 standard deviations among principals serving the smallest number of FRL students (-0.15 SDs and -0.03 SDs, respectively). The difference becomes wider as the percent of FRL students increases, and it becomes 0.54 standard deviations, the widest gap, among principals serving the largest number of FRL students. The average job stress level is 0.36 standard deviations among principals facing the sanctions, whereas it is only -0.18 standard deviations among their colleagues not facing them. The

difference in the differences between the smallest and the largest categories is statistically

significant (p=0.099).

NCLB Sanction and Principals' Turnover Behaviors (Research Question 2)

The previous section finds that the NCLB sanction is associated with greater job stress

and that the job stress level tends to be greater among principals serving a large number of

students from underprivileged backgrounds (i.e., 75% or more). As discussed earlier, when job

stress increases, a person is likely to burn out and then eventually leave his or her workplace. In

this section, I examine whether facing the sanction is associated with higher principal turnover

rates. I estimate Equation 4 through standard logistic regression techniques. Table 9 displays

results. Panel A shows results from the logistic regression model (Model 1) and multinomial

logistic regression model using the three category turnover variable (Model 2); Panel B shows

results from the multinomial logistic regression model using the four category turnover variable

(Model 3). Model 1 reports odds ratios from the logistic regression model; the other models

report relative risk ratios from the multinomial regression model with staying put as a reference

category. All models include school and principal characteristics as well as the number of

schools per district.[63] The middle parts of Appendix Tables 3 and 4 report balance test results for

the binary turnover analysis; the right sides of these tables show balance test results for the three-

category turnover analysis. For brevity, I omit balance test results for the four-category turnover

analysis.

Panel A shows that facing the sanction is associated with higher turnover rates and

transfer rates. For example, the predicted probability that the average principal facing the

---

[63] Because each of these turnover variables has a different number of missing values due to missing information about turnover descriptions, the propensity score matching yielded a slightly different sample for each model.

sanction leaves his or her school at the end of 2007-08 is 23 percent, whereas it is only 16 percent for the average principal facing no sanction. The difference is statistically significant. On the other hand, the difference in transfer rates between the two groups of principals is about two percentage points, which is also statistically significant (4% and 2%, respectively). By four-turnover categories (Panel B), I do not find evidence that the sanction is associated with any of the turnover types, although the sign of the association is in the expected direction. For this analysis, about 13 percent of turnover events were excluded from the sample because the PFS data miss information about districts in which these missing principals worked in 2008-09. This missing information turned the significant association into the insignificant one. Results from this variable may not be generalizable to the target population. For this reason, I do not use the four-category turnover variable as a dependent variable in the subsequent analyses.

Next, like the previous section, I examine moderation effects by principal qualifications and school characteristics in the same way. Table 10 reports estimation results for moderation effects by principal qualifications. I find no evidence that the relationship between the sanction and principal turnover behaviors is moderated by these principal qualifications. None of the coefficients on the interaction terms are statistically significant. Tables 11 and 12 show estimate results for moderation effects by school characteristics. Again, I find little evidence that any school characteristics moderate the relationship between NCLB sanctions and principal turnover.

Influence of State Sanction and Rating Systems

One challenge analyzing the effect of the NCLB sanction is potential influence from state sanction and/or rating systems.[64] For example, according to *Education Week*'s annual *Quality Counts 2007* report, 29 states had accountability systems that assigned ratings to all schools

---

[64] From now on, I use a word, state sanction systems, to mean both state sanction systems and rating systems.

based on their own criteria in the 2006-07 school year (Education Week, 2007). In addition, a

majority of states also used their own sanction systems in addition to the NCLB sanction system.

To the extent that state education agencies used similar performance criteria and gave similar

sanctions to low-performing schools, the influence of the NCLB sanction would be confounded

by that of state sanctions, because schools missing AYP in the NCLB system might also miss

performance standards under the state accountability systems and thus face state sanctions. As a

result, principals might feel pressure from both the NCLB sanction and the state sanctions. My

estimates pick up both influences and get biased upwardly. If state sanction systems are quite

different from the NCLB sanction system in both performance criteria and kinds of sanctions

imposed, then the direction of bias could be either positive or negative.

One strategy to overcome this problem is to use the fact that not all states had their own

sanction system in 2006-07. To identify which states did not have their own sanction systems, I

carefully reviewed *Education Week*'s *Quality Counts* 2007 report.[65] *Education Week* collected

information about each state's sanction system in 2006-07 through surveys sent to state

education policymakers in summer 2006. They vetted state responses and verified the responses

with documentation that included state administration codes, state legislations, state education

agency websites, public and internal reports, meeting minutes, and the like. Table 13 presents a

summary of state sanctions in 2006-07 that *Education Week* collected. It shows that, although

there is a variation across states about which sanctions they used, 13 states (or 10 states in my

sample) did not have their own sanction systems in place in that school year. This means that if a

---

[65] Prior studies of NCLB (Dee & Jacob, 2011; Dee, Jacob, & Schwartz., 2013; Grissom et al., 2014) used the Carnoy and Loeb (2002) index of state accountability strength, which is based on the 1999-2000 school year, and estimated the impact of NCLB using a difference-in-difference estimator. Because I focus on the 2006-07 and 2007-08 school years and there had been substantial changes in state accountability systems between the 1999-2000 and 2006-07 school years, I use detailed state accountability information systematically collected by the Editorial Projects in Education Research Center (EPERC) in 2006-07 for their *Quality Counts* 2007 report. The EPERC is a research division of the Editorial Projects in Education, which publishes *Education Week*.

school in a state without its sanction system faced the NCLB sanction, theoretically, accountability pressure only came from the NCLB sanction system. I use this information to reduce, if not remove, potential bias from state sanction systems. More specifically, I add to Equations 3, 4, and 5 an indicator variable for states that had their own sanction systems in place in 2006-07 to control for the influence of state sanctions, and re-estimate the models.

Table 14 reports estimation results. The inclusion of the state sanction variable does not change the main results. Principals facing the NCLB sanctions feel greater job stress than those not facing them, and they are more likely to turn over or transfer to different schools. No notable changes in the main results suggest that accountability pressure largely came from the NCLB sanction system.

Principal Salary

My main analysis finds evidence that the NCLB sanction is associated with greater job stress, higher turnover rates, and higher transfer rates. These results hold even after removing the influence of state sanction systems. They are consistent with my expectations discussed earlier. One question that arises from these results is whether or not the changes in working conditions are offset by an increase in principals' annual salaries. School districts may increase salaries for principals working at schools facing the sanction to retain them. Because not many districts use principal salary schedules, they have more flexibility in raising principal salaries than teacher salaries.[66] In this section, I investigate this possibility.

After matching principals, I estimate principal salaries as a function of the sanction status and principal and school characteristics through standard multiple regression techniques. I also add an indicator variable for states with their own sanction systems to remove the influence of

---

[66] In 2007-08, about 51 percent of school districts had a salary schedule for principals.

state sanction systems. Principal salaries are adjusted for regional cost differences using the

Comparable Wage Index.[67] For this reason, I do not include district characteristics in the

regression model. Table 15 reports results. It shows that the sanction is not associated with

principal salaries in any model, supporting the claim that principals facing the sanction are not

compensated for changes in working conditions by salary increase.

Job Stress as a Mediator: Causal Mediation Analysis

The two main findings that NCLB sanctions are associated with greater job stress and

higher turnover and transfer rates suggest that job stress may be a mediator between NCLB

sanctions and principal turnover. As discussed earlier, greater job stress generally leads to

burnout, which could induce a person to leave his or her workplace (Leiter & Maslach, 2009).

Given these two findings, it is possible that NCLB-induced job stress made principals burn out

and then eventually leave their schools. As a final analysis, I perform causal mediation analysis.

For this analysis, I first perform propensity score matching with job stress and turnover as

outcome variables. Then, I estimate a causal mediation model where NCLB sanction is a

treatment, job stress is a mediator, and turnover is an outcome. Mediation analysis is traditionally

implemented within the framework of linear structural equation modeling (e.g., MacKinnon,

2008). However, in the current context, the outcome variable is a binary turnover variable, which

needs to be estimated through a logistic or probit regression model. Since a traditional approach

cannot be generalizable to nonlinear models, I estimate the mediation model through a

generalized approach (Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010; Imai,

---

[67] The original Comparable Wage Index is available from the National Center for Education Statistics (Taylor, Glander, & Fowler, 2007). I used the most recent index from http://bush.tamu.edu/research/faculty/Taylor_CWI/.

Keele, Tingley, & Yamamoto, 2009).[68] Panel A of Table 16 reports results. It shows that the average causal mediation effect is 0.00 and statistically indistinguishable from zero, as the 95 percent confidence intervals contain zero. Although the lower bound is very close to zero, suggesting that the estimate would be statistically significant at the 90 percent level, the magnitude of the effect is still small. In fact, the effect only accounts for 6.55 percent of the total effect of NCLB sanctions on principal turnover. The analysis suggests no meaningful mediation effect.

An important assumption for causal mediation analysis is sequential ignorability assumption (see Imai, Keele, & Yamamoto, 2010). It contains two statements of conditional independence. The first statement is that, given observed pretreatment confounders, the treatment assignment is statistically independent of potential outcomes and potential mediators. The second statement is that the mediator is ignorable given observed pretreatment and treatment confounders. While the first part of the assumption could be met in observational studies by including a rich set of pretreatment confounders in the model, the second part is difficult to meet and cannot be directly tested. To understand the degree to which the above finding is robust to potential violation of the section part of the assumption, I perform sensitivity analysis (Imai et al., 2010).[69] In this analysis, I calculate a correlation, ρ, between the error for the mediation model and the error for the outcome model. This correlation increases if there are omitted variables that affect both job stress and principal turnover. This means that when the sequential ignorability assumption is met, ρ is zero. When it is not met, ρ is not equal to zero. The larger the absolute value of ρ, the farther the departure from the assumption. The main purpose of this sensitivity

---

[68] I use user-written Stata commands, *medeff* to estimate the model (Hicks & Tingley, 2011). Because the general approach has not been expanded to incorporate a case where an outcome variable is categorical, I do not estimate the mediation model for a categorical turnover variable.

[69] I use a user-written Stata code, *medsens*, to perform this analysis (Hicks & Tingley, 2011). It is important to note that this code cannot incorporate matching weights. Thus, the analysis needs to be interpreted with care.

analysis is, therefore, to identify the value of ρ, above which the causal mediation effect disappears. The larger the value is, more robust the causal mediation effect is to violation of the assumption.

The first row in Panel B reports that the causal mediation effect goes away when ρ is 0.20 or greater. To visually understand how the effect is sensitive to values of ρ, I plot the average causal mediation effect (ACME) over ρ with the 95 percent confidence interval in Figure 3. It shows that the ACME is relatively stable over different values of ρ. It also shows that, while the ACME becomes zero at ρ equal to 0.20, the 95 percent confidence interval contains the value of zero when approximately $0.15 < \rho < 0.35$. Thus, the finding about the ACME is relatively robust to the violation of the assumption. Returning to Table 16, the last two rows in Panel B report the proportion of unexplained variance explained by unobserved pretreatment confounders and the proportion of the original variance explained by the same unobserved confounders. These proportions are very small (0.04 and 0.03, respectively).

In sum, there do not appear to be strong causal mediation effects by job stress on the relationship between NCLB sanctions and turnover. The effect might be significant at the 10 percent level but the magnitude is very small. The significant relationship between NCLB sanctions and principal turnover largely come from direct effects, although results do not preclude impact of other untested mediators. In addition, the insignificant mediation effect may suggest that turnover could be driven by district administrators. That is, they might have made strategic principal assignment decisions to improve school performance. The current data do not allow for testing, whether principal turnover was voluntary or involuntary.

Discussion and Conclusions

Many studies have investigated the impact of NCLB and its sanction system on students and teachers using large-scale datasets and found that it has increased student performance, especially in mathematics, while its impact on teachers appears to be mixed (e.g., Dee & Jacob, 2011; Grissom et al., 2014; Reback et al., 2014). However, despite the critical roles that school principals play in improving school and student outcomes, prior studies are based on small samples of principals or focused on theoretical arguments (e.g., Brown, 2006; Educational Testing Service, 2008; Lyons & Algozzine, 2006), and there is little empirical evidence on whether NCLB or more specifically, the NCLB sanction is associated with changes in principal behaviors. Because the sanction system was designed to give principals pressure or incentives to change their behaviors in order to avoid severer sanctions, facing early stages of sanctions may have forced principals to work longer hours. If the incentive scheme did not agree with principals' work ethic or professional standards, principals could have had greater job stress. It is possible that these changes in working conditions led principals to leave their schools after facing the sanction. This study provides some sort of new evidence on principals' responses to the NCLB sanction using the Schools and Staffing Survey 2007-08 and detailed school-level assessment and AYP data systematically collected from 45 states.

I find evidence that the NCLB sanction is positively associated with principals' job stress, turnover rates, and transfer rates. Job stress is especially greater among principals serving a large number of underprivileged students and among inexperienced principals. These patterns still hold even after reducing, if not removing, the influence of state sanction systems. I also find that principals facing NCLB sanctions are not compensated for changes in working conditions by salary increase. Finally, I find little evidence that job stress is a mediator for the relationship

183

between NCLB sanctions and principal turnover. This finding is surprising and contrary to what the literature suggests. The significant relationship between NCLB sanctions and principal turnover largely comes from the sanctions' direct effects. Yet, it is still possible that there exist other unobserved mediator(s) that connect NCLB sanctions and principal turnover. Moreover, the null result also suggests that turnover could be driven by district administrators.

These findings have some implications. First, since NCLB, or, more generally, high stakes accountability systems with sanctions, increase principals' job stress level, policymakers may need to provide professional support for principals to reduce their job stress level. It could include assigning additional school support staff to their schools and offering mentoring programs, time management programs, and leadership training programs. Although my study does not pinpoint the source of the job stress, if it directly comes from the inappropriate design of the accountability system, policymakers would need to change the incentive structure such that principals agree with reward/sanction criteria and are motivated to improve their daily practices. Second, if principal turnover was largely driven by principals themselves, not by district administrators, policymakers may need to implement some kind of retention programs to prevent principal turnover to improve the effectiveness of accountability systems. For example, a retention bonus program may work effectively, as low salaries are found to be correlated with higher turnover rates (Akiba & Reicardt, 2004; Baker et al., 2010; Cullen & Mazzeo, 2008). Lastly, given these negative responses to the sanction or, more generally, incentive policies and programs, policymakers would need to consider the design of incentive programs for principals more carefully. Policymakers would need to examine whether performance measures under incentives are appropriate to capture a full dimension of the principalship, as it has multiple dimensions (Dixit, 2002; Podgursky & Springer, 2007). They would also have to assess whether

184

performance targets are set appropriately and differentially (Lazear, 2000; National Research Council, 2011) and whether the size of the incentive is significant enough for principals (Goldhaber, 2007)

This study faces some limitations. First, although the propensity score matching method removes bias resulting from systematic differences in observable characteristics between the treatment and control groups, it cannot correct bias coming from systematic differences in unobservable characteristics. For example, parental and community support for education may be systematically stronger among schools with certain characteristics, which would be associated with both the sanction status and principal responses. Second, the three variables used for principals' working conditions may not be perfect measures of working conditions. It is possible that the NCLB sanction is associated with other measures of working conditions. Third, although I find that principals tend to transfer to different schools when they face sanctions, I do not know what kind of schools they transfer to. Knowing the characteristics of receiving schools would help policymakers understand their turnover behaviors in more depth. Lastly, this analysis is mostly cross-sectional and does not provide insights for long-term effects. While it may not be feasible to track a representative sample of principals over time, pooled cross-sectional data may allow us to answer questions that my analysis could not.

This study will benefit from future work that examines whether NCLB-induced job stress influenced a relationship between the principal and teachers. A current emphasis on principals' instructional leadership roles highlights the importance of a cooperative and supportive relationship between the principal and teachers. Since the sanction is associated with greater job stress among principals, knowing how job stress affected the principal-teacher relationship will be critical to enhance principals' instructional leadership and improve student performance. My

185

study would also benefit from future research that examines the dynamic effect of NCLB sanctions or, more generally, accountability pressure on principal behaviors. My analysis is cross-sectional and does not capture the long-term effect of accountability pressure. District policymakers may find this information useful to make strategic human resource decisions. Finally, my study excluded principals new to their schools. It would be of interest to examine how these principals, especially those hired in already low-performing schools, respond to accountability pressure.

References

Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina*. NBER Working Paper No. 20511. Retrieved from http://www.nber.org/papers/w20511. Cambridge, MA: National Bureau of Economic Research.

Akiba, M. & Reichardt, R. (2004). What predicts the mobility of elementary school leaders? An analysis of longitudinal data in Colorado. *Education Policy Analysis Archives*, *12*(18), 1-20.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*(2), 150-161.

Baker, B., Punswick, E., & Belt, C. (2010). School leadership stability, principal moves, and departures: Evidence from Missouri. *Educational Administration Quarterly, 46*(4), 523–557.

Boser, U., Wilhelm, M., & Hanna, R. (2014). *The power of the Pygmalion effect: Teachers expectations strongly predict college completion*. Washington, DC: Center for American Progress.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals. NBER Working paper No. 17803. Retrieved from http://www.nber.org/papers/w17803. Cambridge, MA: National Bureau of Economic Research.

Brewer, D. J. (1993). Principals and student outcomes: Evidence from U.S. high schools. *Economics of Education Review*, *12*(4), 281-292.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, *163*(12), 1149-1156.

Brooks, K., Adams, S. R., & Morita-Mullaney, T. (2010). Creating inclusive learning communities for ELL students: Transforming school principals' perspectives. *Theory Into Practice*, *49*(2), 145-151.

Brown, P. F. (2006). Preparing principals for today's demands. *Phi Delta Kappan*, *87*(7), 525-526.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24*(4), 305-331.

Chakrabarti, R. (2014). Incentives and responses under No Child Left Behind: Credible threats and the role of competition. *Journal of Public Economics*, *110*, 124-146.

Clark, D., Martorell, P., & Rockoff, J. (2009). *School principals and school performance*. Working paper 38. Washington, DC: The Urban Institute.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, *31*(1), 92-109.

Cullen, J. B. & Mazzeo, M. J. (2008). *Implicit performance awards: An empirical analysis of the labor market for public school administrators*. Working paper. Northwestern University.

Davidson, E., Reback, R., Rockoff, J., & Schwartz, H. L. (2015). Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of NCLB. *Educational Researcher*, *44*(6), 347-358.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418-446.

Dee, T., Jacob, B, & Schwartz, N. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, *35*(2), 252-279.

Dhuey, E., & Smith, J. (2013). *How school principals influence student learning*. Unpublished manuscript.

Diamond, J. B., Randolph, A., & Spillane, J. P. (2004). Teachers' expectations and sense of responsibility for student learning: The importance of race, class, and organizational habitus. *Anthropology and Education Quarterly*, *35*(1), 75-98.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, *37*(4), 696-727.

DuGoff, E. H., Schuler, M., & Stuart, E. A. (2014). Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Services Research*, *49*(1), 284-303.

Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review*, *7*(3), 291-299.

Education Week. (2007). *Quality Counts 2007: From cradle to career*. Bethesda, MD: Editorial Projects in Education.

Educational Testing Service (2008). *Standards, accountability, and flexibility: Americans speak*. Princeton, NJ: Author.

Farley-Ripple, E. N., Solano, P. L., & McDuffie, M. J. (2012). Conceptual and methodological issues in research on school administrators career behavior. *Educational Researcher*, *41*(6), 220-229.

Fruehwirth, J. C., & Traczynski, J. (2013). *Spare the rod? The dynamic effects of failing accountability on schools*. Unpublished manuscript. University of Cambridge.

Gardiner, M. E., & Enomoto, E. K. (2006). Urban school principals and their role as multicultural leaders. *Urban Education*, *41*(6), 560-584.

Gardiner, M. E., Canfield-Davis, K., & Anderson, K. L. (2009). Urban school principals and the 'No Child Left Behind' Act. *The Urban Review*, *41*(2), 141-160.

Goldhaber, D. (2007). *Principal compensation: More research needed on a promising reform*. Washington, D.C.: Center for American Progress.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3-28.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433-444.

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers and their work environment. *Educational Evaluation and Policy Analysis, 36*(4), 417-436.

Gutman, L. M., & McLoyd, V. C. (2000). Parents' management of their children's education within the home, at school, and in the community: An examination of African-American families living in poverty. The *Urban Review*, *32*(1), 1-24.

Hallinger, P. & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*(2), 157-191.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*(3), 234-249.

Hemelt, S. W. (2011). Performance effects of failure to make Adequate Yearly Progress (AYP): Evidence from a regression discontinuity framework. *Economics of Education Review*, *30*, 702-723.

Hicks, R., & Tingley, D. (2011). *Mediation: STATA package for causal mediation analysis*. College Station, TX: StataCorp LP.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309-334.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51-71.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2009). Causal mediation analysis using R. In H. D. Vinod (Ed.), *Advances in Social Science Research Using R* (pp.129-154). New York, NY: Springer.

Johnson, J. V., & Hall, E. M. (1988). Job strain, work place social support, and cardiovascular disease: A cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health*, *78*(10), 1336-1342.

Karasek, R. A. Jr. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, *24*(2), 285-308.

Knapp, M. S., Copland, M. A., Honig, M. I., Plecki, M. L., & Portin, B. S. (2010). *Learning-focused leadership and leadership support: Meaning and practice in urban systems.* Seattle, WA: University of Washington.

Knapp, M. S., & Feldman, S. B. (2012). Managing the intersection of internal and external accountability: Challenge for urban school leadership in the United States. *Journal of Educational Administration*, *50*(5), 666-694.

Koyama, J. (2014). Principals as bricoleurs: Making sense and making do in an era of accountability. *Educational Administration Quarterly*, *50*(2), 279-304.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, *90*(5), 1346-1361.

Lee, J., & Bowen, N. K. (2006). Parent involvement, cultural capital, and the achievement gap among elementary school children. *American Educational Research Journal*, *43*(2), 193-218.

Leiter, M. P., & Maslach, C. (2009). Nurse turnover: the mediating role of burnout. *Journal of Nursing Management*, *17*(3), 331-339.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning: A review of research for the Learning from Leadership Project*. New York, NY: Wallace Foundation.

Li, D. (2012). *School accountability and principal mobility: How No Child Left Behind affects the allocation of school leaders*. Working paper. Cambridge, MA: Harvard University.

Louis, K. S., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). *Investigating the links to improved student learning: Final report of research findings*. Minneapolis, MN: University of Minnesota and Toronto, ON: University of Toronto.

Lyons, J. E., & Algozzine, B. (2006). Perceptions of the impact of accountability on the role of principals. *Education Policy Analysis Archives*, *14*(16), 1-16.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Routledge.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*, 397-422.

McGhee, M. W., & Nelson, S. W. (2005). Sacrificing leaders, villainizing leadership: How educational accountability policies impair school leadership. The *Phi Delta Kappan*, *86*(5), 367-372.

National Research Council. (2011). *Incentives and Test-Based Accountability in Public Education*. Committee on Incentives and Test-Based Accountability in Public Education, Michael Hout and Stuart W. Elliott, *Editors*. Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, *26*(4), 909-949.

Portin, B., Schneider, P., DeArmond, M., & Gundlach, L. (2003). *Making sense of leading schools: A study of the school principalship*. Seattle, WA: University of Washington.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, *6*(3), 207-241.

Ridgeway, G., Kovalchik, S. A., Griffin, B. A., & Kabeto, M. U. (2015). Propensity score analysis with survey weighted data. *Journal of Causal Inference*, *3*(2), 237-249.

Rogoff, B. (2003). *The cultural nature of human development*. New York, NY: Oxford University Press.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169-188.

Salazar, K. (2007). *Senator Salazar's No Child Left Behind survey results*. Retrieved on September 17, 2015, from http://salazar.senate.gov/other/nclbresults.html#who.

Schnall, P. L., Lansbergis, P. A., & Baker, D. (1994). Job strain and cardiovascular disease. *Annual Review of Public Health*, *15*, 381-411.

Schuler, R. S. (1979). Managing stress means managing time. *Personnel Journal (Pre-1986)*, *58*(12), 851-854.

Shipps, D. (2012). Empowered or beleaguered? Principals' accountability under New York City's diverse provider regime. *Education Policy Analysis Archives*, *20*(1). Retrieved on June 4, 2016, from http://epaa.asu.edu/ojs/article/view/892.

Solano, P. L., McDuffie, M. J., & Farley-Ripple, E. N., & Bruton, J. (2010). *Principal retention in the state of Delaware.* Newark, DE: University of Delaware, Center for Community Research and Service.

Spillane, J. P. (2005). Distributed leadership. *The Educational Forum*, *69*(2), 143-150.

Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, *16*(5), 731-762.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, *30*(3), 23-28.

Spillane, J. P., & Kenney, A. W. (2012). School administration in a changing education sector: the US experience. *Journal of Educational Administration*, *50*(5), 541-561.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1-21.

Taylor, L. L., Glander, M., & Fowler, W. J. (2007). *Documentation for the NCES Comparable Wage Index Data File, 2005* (EFSC 2007-397). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

United States Department of Education. (2002). *No Child Left Behind: A desktop reference 2002*. Washington, D.C.: United States Department of Education.

Van der Doef, M., & Maes, S. (1999). The job demand-control (-support) model and psychological well-being: A review of 20 years of empirical research. *Work & Stress*, *13*(2), 87-114.

Wallace Foundation. (2013). *The school principal as leader: Guiding schools to better teaching and learning*. Washington, DC: Author.

Wang, X. (2015). Pathway to a Baccalaureate in STEM fields: Are community college a viable route and does early STEM momentum matter? *Educational Evaluation and Policy Analysis*, *37*(3), 376-393.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced Leadership: What 30 years of research tells us about the effect of leadership on student achievement* (Working paper). Aurora, CO: Mid-continent Research for Education and Learning.

White, B. R., & Agarwal, P. K. (2011). *The principal report: The State of School Leadership in Illinois*. Edwardsville, IL: Illinois Education Research Council, Southern Illinois University Edwardsville.

Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioural health of children and youth. *American Psychologist*, *67*(4), 272-284.

Zanutto, E. L. (2006). A comparison of propensity score and linear regression analysis of complex survey data. *Journal of Data Science*, *4*(1), 67-91.
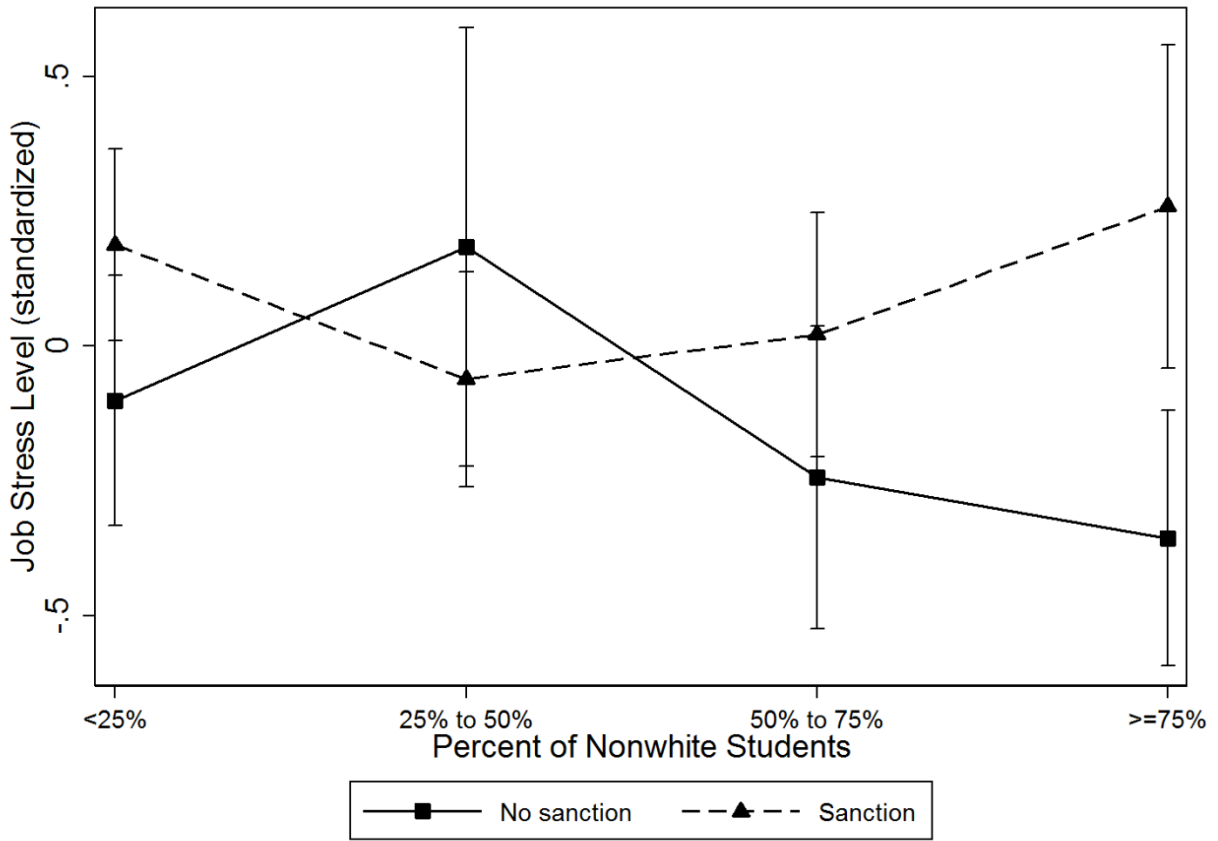
Figure 1: Predicted job stress level by sanction status over percent of nonwhite students
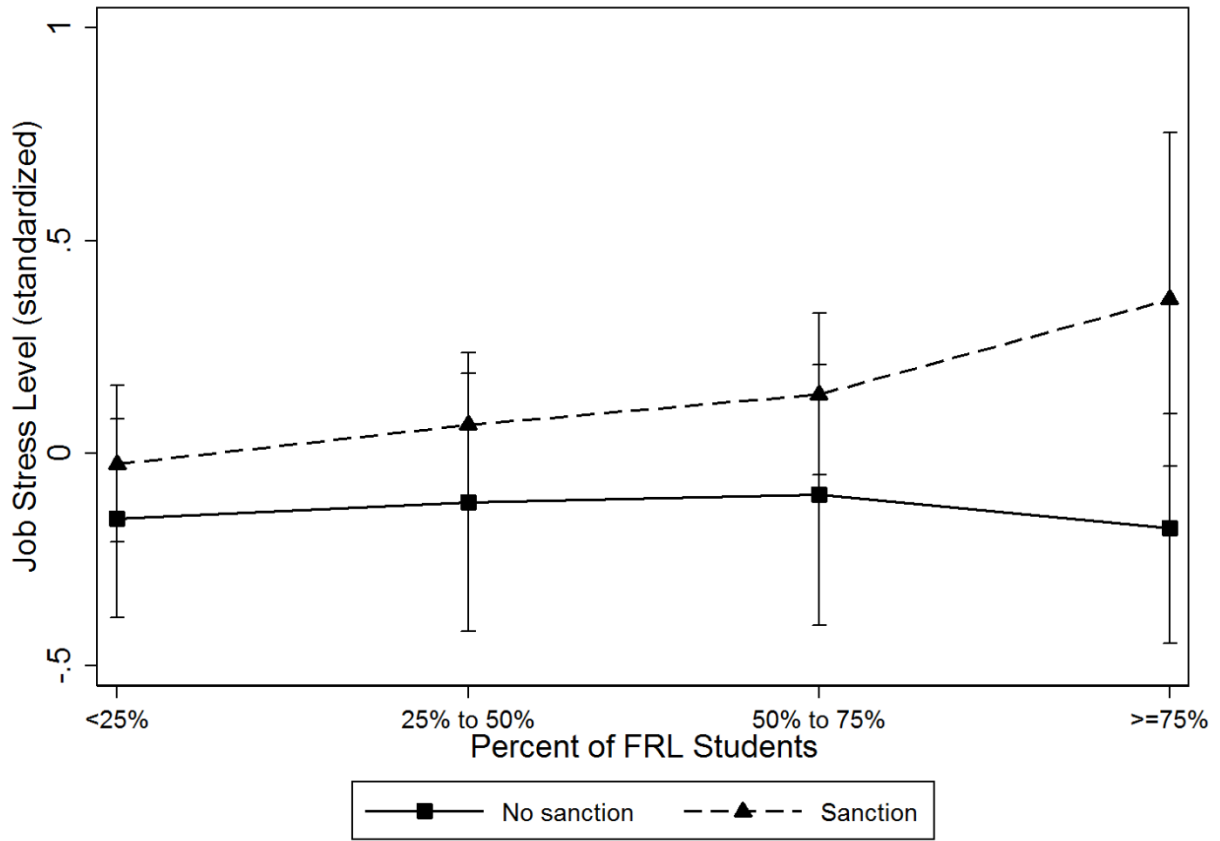
Figure 2: Predicted job stress level by sanction status over percent of FRL students
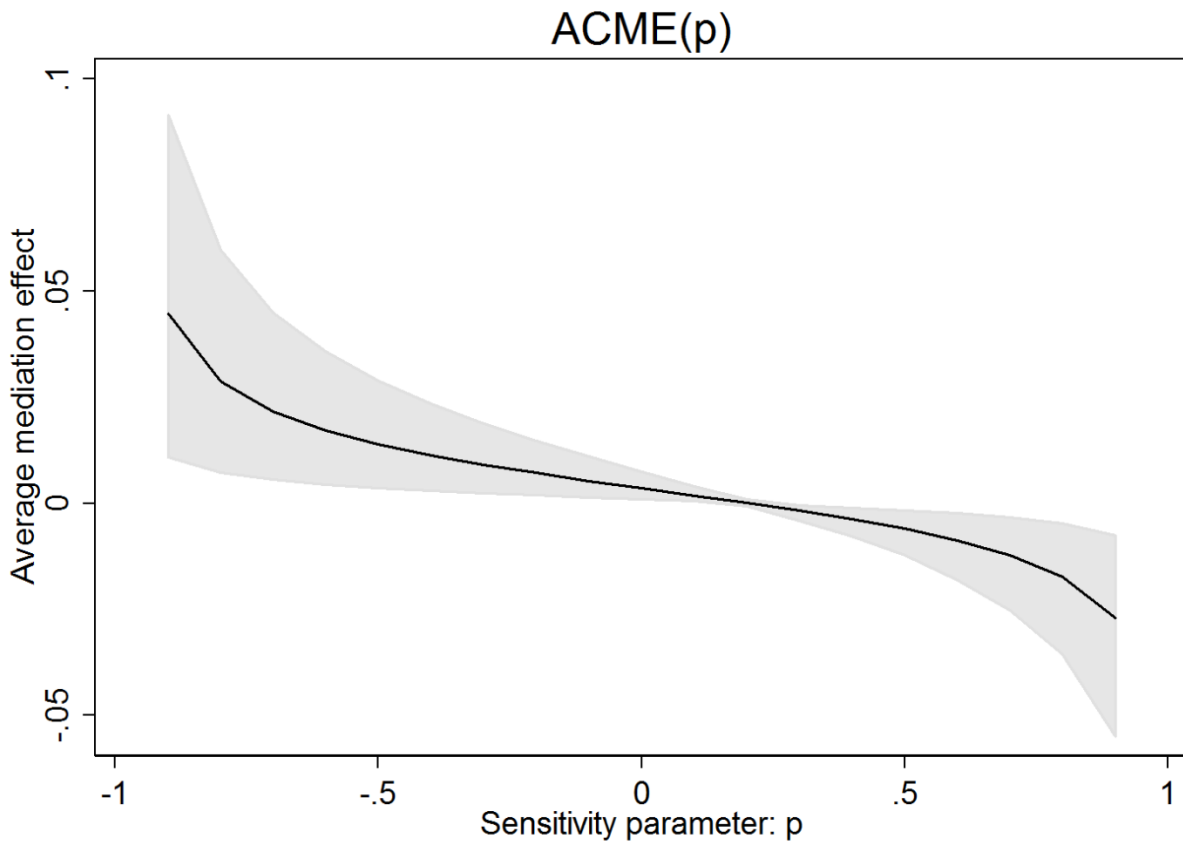
Figure 3: Average causal mediation effect over different values of ρ

Table 1: NCLB sanction system

Number of *consecutive* years that school missed AYP in same subject in at the end of year *t-1*	Sanction imposed in the *following* school year (year *t*)
0	No sanction imposed.
1	No sanction imposed but the failure is publicly announced.
2	School Improvement Year 1 (SIY 1). District must offer Public School Choice (PSC) with transportation to all students to transfer to another public school or charter school within the same district which has not been identified for "in-need-of improvement." It also needs to provide technical assistance.
3	School Improvement Year 2 (SIY 2). District must offer supplemental educational services (SES) to students from low-income families in addition to PSC and technical assistance.
4	School Improvement Year 3 (SIY 3). District must take at least one of the following corrective actions on failing schools: (1) replace school staff, (2) implement a new curriculum (with appropriate professional development), (3) decrease management authority at the school level, (4) appoint an outside expert to advise the school, (5) extend the school day or year, and (6) reorganize the school internally. District continues to offer PSC, SES, and technical assistance.
5	School Improvement Year 4 (SIY 4). District must initiate plans to fundamentally restructure failing schools. It continues to offer PSC, SES, and technical assistance.
6	School Improvement Year 5 (SIY 5). District must implement the school restructuring plan. It includes one or more of the following actions: (1) reopen the school as a public or charter school, (2) replace all or most of the staff who are responsible for the failure to make AYP, (3) enter into a contract with an entity, such as private management company to operate the school, (4) turn the operation of the school over to the state department of education, and (5) other major restructuring. District continues to offer PSC, SES, and technical assistance.
7 and beyond	School remains in restructuring until it meets AYP for two consecutive years.

Notes: If a district does not have PSC options because all schools are identified for "in-need-of improvement," it offers SES instead. Data source: US Department of Education

Table 2: AYP determination rules in 2006-07

	Use of Confidence Interval (CI)			Safe Harbor			Uniform Averaging	Growth Model
	Applied	Confidence Level (%)	Applied to Performance	Applied	CI Applied	Confidence Level	Applied	Applied
Alabama	Yes	99	Yes	Yes			Yes	
Alaska	Yes	99		Yes	Yes	75%		Yes
Arizona	Yes	99	Yes	Yes				Yes
Arkansas	unknown	unknown	unknown	Yes	Yes	75%	Yes	Yes
California	Yes	99	Yes	Yes	Yes	75%	Yes	
Colorado	Yes	95	Yes	Yes				
Connecticut	Yes	99	Yes	Yes	Yes	75%		
Delaware	Yes	98	Yes	Yes	Yes	75%	Yes	Yes
Florida	Yes	NA	Yes	Yes			Yes	Yes
Georgia	Yes	95	Yes	Yes			Yes	
Hawaii	Yes	NA	Yes	Yes			Yes	
Idaho				Yes			Yes	
Illinois	Yes	95	Yes	Yes	Yes	75%		
Indiana	Yes	99	unknown	Yes	Yes	75%		
Iowa	Yes	98	Yes	Yes			Yes	Yes
Kansas	Yes	99		Yes	Yes	75%	Yes	
Kentucky	Yes	99	Yes	Yes			Yes	
Louisiana	Yes	99		Yes	Yes	99%		
Maine	Yes	95	Yes	Yes	Yes	75%		
Maryland	Yes	95		Yes	Yes	unknown		
Massachusetts	Yes	NA		Yes				
Michigan	Yes	95	Yes	Yes			Yes	
Minnesota	Yes	95	unknown	Yes			Yes	
Mississippi	Yes	99	Yes	Yes				
Missouri	Yes	99	Yes	Yes	Yes	75%	Yes	

State								
Montana	Yes	95	Yes	Yes	Yes	75%		
Nebraska	Yes	99		Yes	Yes	75%		
Nevada	Yes	95	Yes	Yes	Yes	75%	Yes	
New Hampshire	Yes	99	unknown	Yes	unknown	unknown		
New Jersey	Yes	95	Yes	Yes	Yes	75%		
New Mexico	Yes	99		Yes				
New York	Yes	90		Yes			Yes	
North Carolina	Yes	95	Yes	Yes				Yes
North Dakota	Yes	99	unknown	Yes			Yes	
Ohio				Yes			Yes	Yes
Oklahoma	Yes	95	Yes	Yes	Yes	75%		
Oregon	Yes	99	Yes	Yes			Yes	
Pennsylvania	Yes	95	Yes	Yes	Yes	75%	Yes	
Rhode Island	Yes	95	Yes	Yes			Yes	
South Carolina	Yes	95	Yes	Yes			Yes	
South Dakota	Yes	99	Yes	Yes	Yes	75%	Yes	
Tennessee	Yes	95		Yes				Yes
Texas	Yes	95	Yes	Yes			Yes	
Utah	Yes	99	Yes	Yes	Yes	75%		
Vermont	Yes	99		Yes				
Virginia	Yes	NA	unknown	Yes			Yes	
Washington	Yes	99	Yes	Yes			Yes	
West Virginia	Yes	99		Yes			Yes	
Wisconsin	Yes	99	Yes	Yes	Yes	75%		
Wyoming	Yes	95	Yes	Yes	Yes	75%	Yes	
Total	47		31	50	22		28	9

Notes: A blank cell means No. unknown means that I could not find information. NA means not applicable. Note that, for these proficiency adjustment methods, some states applied them only to small schools defined by their criteria.

Table 3: Descriptive statistics

| Variable | N | All Schools | | | | By AYP Status (2006-07) | |
		Mean	SD	Min	Max	Made AYP	Missed AYP
Principal characteristics							
Age	3200	49.4	8.35	24	75	49.4	49.5
Female	3200	0.52		0	1	0.53	0.47**
American Indian	3200	0.01		0	1	0.01	0.00
Asian	3200	0.01		0	1	0.01	0.01
Black	3200	0.09		0	1	0.08	0.15***
Hispanic	3200	0.05		0	1	0.05	0.05
White	3200	0.85		0	1	0.87	0.80***
Years of experience as principal	3200	7.46	6.67	0	43	7.55	7.10
Years in current school	3200	4.15	4.71	0	37	4.27	3.67**
Education specialist or doctoral degree	3200	0.40		0	1	0.40	0.41
School characteristics							
Percent nonwhite students	3200	33.57	30.63	0	100	31.83	40.79***
Percent FRL students	3090	39.96	25.88	0	99.78	38.87	44.57***
Elementary school	3200	0.63		0	1	0.69	0.38***
Middle school	3200	0.20		0	1	0.16	0.38***
High school	3200	0.15		0	1	0.13	0.21***
Combination school	3200	0.02		0	1	0.02	0.04***
School enrollment size	3200	587	440	0	5324	541	776***
Urban	3200	0.20		0	1	0.19	0.26***
Suburban	3200	0.33		0	1	0.33	0.34
Town	3200	0.15		0	1	0.15	0.15
Rural	3200	0.31		0	1	0.33	0.24***
Title I	3110	0.62		0	1	0.63	0.61
District characteristics							
Number of schools	3110	36.90	86.14	1	829	33.23	52.50***

Notes: All variables are based on 2006-07. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. A total of 45 states are included. Survey weights are used. Oklahoma has assessment data but because the scale of their AMO is completely different from that of the other states, I dropped the state from the sample. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Principals' working conditions and turnover behaviors by AYP status

	All Schools					By AYP Status (2006-07)	
	N	Mean	SD	Min	Max	Made AYP	Missed AYP
Working conditions							
Total hours spent on school-related activities per week	3200	58.87	10.88	1.00	150.00	58.73	59.47
Total hours spent on interactions with students per week	3200	20.69	12.84	0.00	150.00	20.27	22.45***
Job stress (standardized)	3200	-0.04	0.97	-1.04	3.80	-0.06	0.08**
Turnover							
Leaver	3190	0.21	0.41		1	0.20	0.25*
Mover	3160	0.07	0.25		1	0.06	0.09*
Within-district mover	3140	0.04	0.19		1	0.03	0.06*
Across-district mover	3140	0.03	0.16		1	0.03	0.03
Exit	3160	0.13	0.33		1	0.13	0.14

Notes: All outcome variables are based on 2007-08. Positive values mean greater job stress. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. A total of 45 states are included. Sample sizes are slightly smaller for the mover and exit variables than those in the leaver variable because some principals did not provide adequate information about their new work places. Survey weights are used. Oklahoma has assessment data but because the scale of their AMO is completely different from that of the other states, I dropped the state from the sample. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: NCLB sanction and principals' working conditions

	Total Hours Spent on School Activities Per Week	Total Hours Spent on Interactions with Students Per Week	Job Stress (Standardized)
	Model 1	Model 2	Model 3
Facing sanction	0.31	1.44	0.24***
	(0.86)	(0.97)	(0.08)
Constant	59.27***	18.37***	0.48
	(3.51)	(3.85)	(0.30)
Observations	1210	1210	1210
Adjusted R-squared	0.08	0.06	0.06

Notes: A product of survey weights and matching weights is used. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. A total of 45 states are included. All models include principal and school characteristics. Standard errors are clustered at the district level and reported in parentheses. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Moderation effects by principal qualifications – working conditions

	Total Work Hours Spent on School Activities Per Week			Total Hours Spent on Interactions with Students Per Week			Job Stress (Standardized)		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Facing sanction	0.06	-1.32	-1.20	2.79**	2.10	0.89	0.12	0.26**	0.31**
	(1.12)	(1.44)	(1.64)	(1.24)	(1.38)	(1.52)	(0.12)	(0.12)	(0.12)
Sanction x ed. specialist/doctorate	0.65			-3.50*			0.31**		
	(1.76)			(2.02)			(0.15)		
Sanction x prin exp		0.25			-0.10			-0.00	
		(0.15)			(0.15)			(0.01)	
Sanction x prin exp (4-6 yrs)			2.90			2.17			-0.02
			(2.19)			(2.59)			(0.23)
Sanction x prin exp (7-9 yrs)			-1.22			-0.43			-0.25
			(2.48)			(3.31)			(0.27)
Sanction x prin exp (>=10 yrs)			4.63**			0.35			-0.06
			(2.20)			(2.55)			(0.19)
Constant	59.36***	59.79***	60.45***	17.90***	18.16***	19.77***	0.52*	0.47	0.43
	(3.61)	(3.57)	(3.71)	(3.89)	(3.93)	(3.89)	(0.30)	(0.30)	(0.32)
Observations	1210	1210	1210	1210	1210	1210	1210	1210	1210
Adjusted R-squared	0.08	0.09	0.09	0.06	0.06	0.06	0.06	0.05	0.06

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. All models include principal and school characteristics. Standard errors are clustered at the district level and reported in parentheses. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Moderation effects by Title I school status and school level – working conditions

	Total Work Hours Spent on School Activities Per Week		Total Hours Spent on Interactions with Students Per Week		Job Stress (Standardized)	
	Model 1	Model 2	Model 4	Model 5	Model 7	Model 8
Facing sanction	-0.53	-1.16	1.17	-0.63	0.23**	0.19**
	(1.12)	(1.26)	(1.53)	(1.43)	(0.10)	(0.09)
Sanction x Title I status	1.46		0.48		0.02	
	(1.83)		(2.09)		(0.16)	
Sanction x high school		-0.25		0.04		-0.06
		(2.13)		(2.66)		(0.16)
Constant	59.60***	60.47***	18.48***	16.03***	0.48	0.79***
	(3.48)	(2.98)	(3.90)	(3.69)	(0.30)	(0.30)
Observations	1210	1160	1210	1160	1210	1160
Adjusted R-squared	0.08	0.05	0.06	0.02	0.05	0.05

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. All models include principal and school characteristics. Models 1, 4, and 7 are identical to those in Table 4. Standard errors are clustered at the district level and reported in parentheses. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Moderation effect by percent of nonwhite students – working conditions

Panel A: Nonwhite students	Total Work Hours Spent on School Activities Per Week		Total Hours Spent on Interactions with Students Per Week		Job Stress (Standardized)	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Facing sanction	1.25	1.55	2.40	2.08	0.14	0.29**
	(1.38)	(1.06)	(1.61)	(1.51)	(0.13)	(0.12)
Sanction x percent nonwhite students	-0.02		-0.02		0.00	
	(0.03)		(0.03)		(0.00)	
Sanction x Percent nonwhite students (25% to 50%)		-3.22		1.83		-0.53**
		(3.05)		(2.37)		(0.26)
Sanction x Percent nonwhite students (50% to 75%)		-0.16		-2.36		-0.02
		(2.24)		(3.03)		(0.22)
Sanction x Percent nonwhite students (75% or more)		-2.54		-2.78		0.33
		(2.21)		(2.83)		(0.20)
Constant	58.83***	58.93***	17.92***	18.34***	0.53*	0.44
	(3.63)	(3.33)	(3.92)	(3.90)	(0.30)	(0.29)
Observations	1210	1210	1210	1210	1210	1210
Adjusted R-squared	0.08	0.08	0.06	0.06	0.06	0.07

Panel B: Low-income students	Total Work Hours Spent on School Activities Per Week		Total Hours Spent on Interactions with Students Per Week		Job Stress (Standardized)	
	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Facing sanction	0.32	1.10	0.96	0.75	0.09	0.13
	(1.82)	(1.36)	(1.95)	(1.81)	(0.15)	(0.13)
Sanction x percent FRL students	-0.00		0.01		0.00	
	(0.03)		(0.04)		(0.00)	
Sanction x Percent FRL students (25% to 50%)		-1.27		1.04		0.05
		(2.60)		(2.54)		(0.21)

Sanction x Percent FRL students (50% to 75%)		-0.90		0.09		0.11
		(1.79)		(2.57)		(0.23)
Sanction x Percent FRL students (75% or more)		-0.84		1.28		0.41*
		(2.80)		(3.24)		(0.25)
Constant	59.27***	58.55***	18.64***	18.00***	0.56*	0.57*
	(3.68)	(3.03)	(3.99)	(3.82)	(0.31)	(0.32)
Observations	1210	1210	1210	1210	1210	1210
Adjusted R-squared	0.08	0.08	0.06	0.06	0.06	0.05

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. All models include principal and school characteristics. Standard errors are clustered at the district level and reported in parentheses. Models 1, 3, 5, 7, 9, and 11 are identical to those in Table 4. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: NCLB sanction and principal turnover

Panel A: Results from logistic/multinomial regression models (three categories)	Model 1	Model 2	
	Turnover	Transfer	Exit
Facing sanction	1.63**	2.50**	1.20
	(2.37)	(2.24)	(0.75)
Observations	1200	1200	
Pseudo R-squared	0.03	0.06	
Log Likelihood	-7596.06	-9157.70	

Panel A: Results from multinomial regression models (four categories)	Model 3		
	Within-district Transfer	Across-district Transfer	Exit
Facing sanction	1.56	1.17	1.25
	(1.05)	(0.33)	(0.91)
Observations	1180		
Pseudo R-squared	0.10		
Log Likelihood	-9504.97		

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. All models include principal and school characteristics. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. For each model, a propensity score matching was performed, which yielded a different matched sample for each model. Odds ratios are reported in Model 1. Relative risk ratios are reported in the other models. *z* statistics are reported in parentheses. Standard errors are clustered at the district level. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Moderation effect by principal qualifications – principal turnover

	Model 1	Model 2		Model 3	Model 4		Model 5	Model 6	
	Turnover	Transfer	Exit	Turnover	Transfer	Exit	Turnover	Transfer	Exit
Facing sanction	2.00**	2.74*	1.36	1.54	1.62	1.02	1.70	2.83**	1.08
	(2.56)	(1.94)	(0.94)	(1.31)	(0.80)	(0.06)	(1.40)	(2.07)	(0.21)
Sanction x ed. specialist/doctorate	0.57	0.76	0.71						
	(-1.35)	(-0.37)	(-0.71)						
Sanction x prin exp				1.01	1.10	1.02			
				(0.26)	(1.18)	(0.50)			
Sanction x prin exp (4-6 yrs)							1.05	0.90	0.63
							(0.08)	(-0.17)	(-0.98)
Sanction x prin exp (7-9 yrs)							0.64	0.51	0.82
							(-0.72)	(-1.16)	(-0.40)
Sanction x prin exp (>=10 yrs)							1.02	0.68	1.50
							(0.03)	(-0.69)	(1.05)
Observations	1200	1200		1200	1200		1200	1200	
Pseudo R-squared	0.04	0.06		0.03	0.06		0.03	0.05	
Log Likelihood	-7575.61	-9150.71		-7595.23	-9136.86		-7578.90	-9288.50	

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. All models include principal and school characteristics. Odds ratios are reported in Models 1, 3, and 5. Relative risk ratios are reported in the other models. Standard errors are clustered at the district level. z statistics are reported in parentheses. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p <$ 0.10, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Moderation effect by Title I school status and school level – principal turnover

	Model 1	Model 2		Model 3	Model 4	
	Turnover	Transfer	Exit	Turnover	Transfer	Exit
Facing sanction	1.78*	2.61	1.98**	1.43	1.82	0.98
	(1.91)	(1.57)	(1.99)	(1.28)	(0.73)	(0.33)
Sanction x Title I status	0.88	0.93	0.46			
	(-0.33)	(-0.09)	(-1.60)			
Sanction x high school				1.03	0.42	2.06
				(0.06)	(0.29)	(1.03)
Observations	1200	1200		1130	1140	
Pseudo R-squared	0.03	0.06		0.05	0.05	
Log Likelihood	-7594.87	-9129.91		-7503.07	-9162.97	

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. All models include principal and school characteristics. Odds ratios are reported in Model 1. Relative risk ratios are reported in the other models. Models 1, 3, and 5 are identical to those in Table 6. z statistics are reported in parentheses. Standard errors are clustered at the district level. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Moderation effect by percent of nonwhite students and percent of FRL students – principal turnover

Panel A: Nonwhite students	Model 1			Model 3		
	Model 1	Model 2		Model 3	Model 4	
	Turnover	Transfer	Exit	Turnover	Transfer	Exit
Facing sanction	1.53	2.88	1.86	1.53	1.72	1.19
	(1.25)	(1.57)	(1.47)	(1.35)	(1.04)	(0.51)
Sanction x percent nonwhite students	1.00	1.00	0.99			
	(0.26)	(-0.23)	(-1.31)			
Sanction x Percent nonwhite students (25% to 50%)				0.97	2.54*	0.93
				(-0.06)	(1.89)	(-0.17)
Sanction x Percent nonwhite students (50% to 75%)				0.97	1.01	0.86
				(-0.06)	(0.02)	(-0.34)
Sanction x Percent nonwhtie students (75% or more)				1.36	0.92	1.12
				(0.51)	(-0.12)	(0.21)
Observations	1200	1200		1200	1200	
Pseudo R-squared	0.03	0.06		0.04	0.03	
Log Likelihood	-7595.19	-9131.68		-7558.65	-9427.45	

Panel B: Low-income student	Model 5			Model 7		
	Model 5	Model 6		Model 7	Model 8	
	Turnover	Transfer	Exit	Turnover	Transfer	Exit
Facing sanction	1.55	3.19	2.67*	2.12*	3.76**	1.39
	(1.00)	(1.24)	(1.91)	(1.90)	(2.54)	(0.91)
Sanction x percent FRL students	1.00	0.99	0.98*			
	(0.13)	(-0.28)	(-1.76)			
Sanction x Percent FRL students (25% to 50%)				0.55	0.26**	0.66
				(-1.11)	(-2.41)	(-0.96)
Sanction x Percent FRL students (50% to 75%)				0.87	0.74	0.88
				(-0.24)	(-0.57)	(-0.32)
Sanction x Percent FRL students (75% or more)				0.96	0.94	1.12

				(-0.06)	(-0.08)	(0.22)
Observations		1200	1200	1200	1200	
Pseudo R-squared		0.03	0.06	0.04	0.03	
Log Likelihood		-7595.81	-9117.25	-7525.23	-9395.67	

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. All models include principal and school characteristics. Odds ratios are reported in Models 1, 2, 5, and 6. Relative risk ratios are reported in the other models. Models 1, 3, 5, and 7 are identical to those in Table 5. z statistics are reported in parentheses. Standard errors are clustered at the district level. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Own Rating system	Sanctions authorized under state policy include:									
		Withholding funds	Offering supplemental services	School choice	Implementing new curriculum	Reconstitution	Restructuring	School closure	Reopening school as charter	Management turnover to private firm	Management turnover to state agency
Alabama			Yes		Yes	Yes	Yes			Yes	Yes
Alaska											
Arizona	Yes		Yes			Yes				Yes	
Arkansas	Yes		Yes	Yes		Yes	Yes	Yes			
California	Yes			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colorado	Yes				Yes	Yes	Yes	Yes	Yes	Yes	
Connecticut											
Delaware	Yes		Yes		Yes	Yes	Yes		Yes	Yes	
Florida	Yes		Yes	Yes	Yes	Yes	Yes			Yes	
Georgia	Yes		Yes	Yes	Yes	Yes	Yes		Yes	Yes	
Hawaii				Yes			Yes		Yes		Yes
Idaho			Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Illinois	Yes	Yes				Yes	Yes	Yes	Yes	Yes	Yes
Indiana	Yes		Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes
Iowa							Yes				
Kansas	Yes										
Kentucky	Yes			Yes	Yes	Yes	Yes				Yes
Louisiana	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Maine											
Maryland		Yes			Yes	Yes	Yes	Yes	Yes	Yes	Yes
Massachusetts	Yes		Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Michigan	Yes			Yes			Yes		Yes		

State											
Minnesota	Yes										
Mississippi	Yes		Yes			Yes					Yes
Missouri											
Montana											
Nebraska											
Nevada											
New Hampshire											
New Jersey											
New Mexico	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
New York	Yes		Yes		Yes	Yes	Yes	Yes	Yes		
North Carolina	Yes					Yes				Yes	
North Dakota			Yes	Yes	Yes	Yes	Yes				
Ohio	Yes	Yes		Yes	Yes	Yes	Yes		Yes	Yes	Yes
Oklahoma	Yes			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Oregon	Yes										
Pennsylvania											
Rhode Island	Yes		Yes			Yes	Yes	Yes	Yes		Yes
South Carolina	Yes		Yes			Yes	Yes				Yes
South Dakota											
Tennessee	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Texas	Yes			Yes		Yes	Yes	Yes		Yes	Yes
Utah	Yes										
Vermont					Yes		Yes	Yes			Yes
Virginia	Yes				Yes	Yes	Yes	Yes			
Washington											
West Virginia	Yes		Yes	Yes	Yes	Yes	Yes				Yes
Wisconsin											
Wyoming				Yes	Yes	Yes	Yes				
Total	29	5	18	18	23	30	29	15	16	18	20

Notes: Colorado schools receiving an academic-performance rating of "unsatisfactory" may enter into voluntary restructuring. Georgia does not assign ratings to low-performing schools other than AYP designations. It labels high-performing schools based on a combination of AYP and state-developed performance indicators. The Illinois School Code authorizes the state board of education to change the recognition status of a school or school district to "nonrecognized." The Maryland General Assembly approved a one-year moratorium that prohibits the state board of education from removing a public school from the direct control of the Baltimore City Board of School Commissioners or implementing a major restructuring of a governance arrangement of a public school in the Baltimore city school system. Source: Quality Counts 2007: From Cradle to Career. Editorial Projects in Education Research Center, 2007

Table 14: NCLB sanction and principal behaviors with state sanction systems controlled

Panel A: Working conditions	Total Hours Spent on School Activities Per Week	Total Hours Spent on Interactions with Students Per Week	Job Stress (Standardized)
	Model 1	Model 2	Model 3
Facing sanction	0.29	1.45	0.23***
	(0.85)	(0.97)	(0.08)
Constant	58.94***	18.41***	0.40
	(3.49)	(3.85)	(0.29)
Observations	1210	1210	1210
Adjusted R-squared	0.08	0.06	0.06

Panel B: Turnover (three categories)	Model 4	Model 5	
	Turnover	Transfer	Exit
Facing sanction	1.64**	2.55**	1.24
	(2.39)	(2.32)	(0.93)
Observations	1200	1200	
Pseudo R-squared	0.03	0.06	
Log Likelihood	-7592.53	-9134.00	

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. All models include principal and school characteristics. Odds ratios are reported in Model 4. Relative risk ratios are reported in Models 5 and 6. Standard errors are clustered at the district level and reported in parentheses in Panel A. z statistics are reported in parentheses in Panel B. Standard errors are also clustered at the district level. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: NCLB sanction and principal salary

	Model 1	Model 2
Facing sanction	552.35	567.70
	(1081.68)	(1069.10)
Constant	70267.42***	73934.45***
	(4534.91)	(4321.25)
Control for state sanction systems	No	Yes
Observations	1190	1190
Adjusted R-squared	0.29	0.30

Notes: A product of survey weights and matching weights is used. A total of 45 states are included. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. Principals without school-level assessment data in 2006-07 are excluded. All models include principal and school characteristics. Salaries are adjusted for regional cost differences using the Comparable Wage Index. Standard errors are clustered at the district level and reported in parentheses. Sample sizes are rounded to the nearest ten due to NCES non-disclosure policies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Causal mediation analysis

Panel A: Mediation analysis	Estimate	95% Confidence Interval	
Average mediation effect	0.00	0.00	0.01
Average direct effect	0.07	0.01	0.13
Percent of total effect mediated	6.55	3.57	32.11

Panel B: Sensitivity analysis	Estimate
ρ at which ACME = 0	0.20
$R^2_M * R^2_Y*$ at which ACME = 0	0.04
$R^2_M \sim R^2_Y \sim$ at which ACME = 0	0.03

Notes: The causal mediation analysis and the sensitivity analysis are performed using user-written Stata codes, *medeff* and *medsens* (Hicks & Tingley, 2011). ACME denotes average causal mediation effect. For the sensitivity analysis, matching weights are not incorporated.

Appendix Table 1: Factor analysis of principals' job stress

Job Stress	Factor loadings	Average scale
		Alpha = 0.74
The stress and disappointments involved in serving as principal at this school aren't really worth it.	0.80	1.64
If I could get a higher paying job, I'd leave education as soon as possible.	0.74	1.77
I think about transferring to another school.	0.72	1.66
I don't seem to have as much enthusiasm now as I did when I began my career as a principal.	0.81	1.80
I think about staying home from school because I'm just too tired to go.	0.74	1.40

Items are on a 1-4 Likert scale. The higher values mean greater job stress. One-factor solution employs varimax rotation. An eigenvalue for this factor is 3.30.

Appendix Table 2: Comparison between the analytic and full samples

Variable	States with Assessment Data (45 states)		All states (50 states)		Comparison	
	N	Mean	N	Mean	Difference	p value
Principal characteristics						
Age	3200	49.4	5310	49.46	-0.1	0.82
Female	3200	0.52	5310	0.49	0.03	0.04
American Indian	3200	0.01	5310	0.01	0.00	0.01
Asian	3200	0.01	5310	0.01	0.00	0.32
Black	3200	0.09	5310	0.09	0.00	0.87
Hispanic	3200	0.05	5310	0.04	0.01	0.45
White	3200	0.85	5310	0.86	0.01	0.45
Years of experience as principal	3200	7.46	5310	7.51	-0.05	0.78
Years in current school	3200	4.15	5310	4.20	-0.05	0.71
Education specialist or doctoral degree	3200	0.40	5310	0.39	0.01	0.30
School characteristics						
Percent nonwhite students	3200	33.57	5030	32.58	0.99	0.24
Percent FRL students	3090	39.96	4850	40.42	-0.46	0.51
Elementary school	3200	0.63	5100	0.57	0.06	0.00
Middle school	3200	0.20	5100	0.17	0.03	0.00
High school	3200	0.15	5100	0.21	-0.06	0.00
Combination school	3200	0.02	5100	0.05	-0.03	0.00
School enrollment size	3200	587	5080	533	54	0.00
Urban	3200	0.20	5160	0.20	0.00	0.75
Suburban	3200	0.33	5160	0.30	0.03	0.01
Town	3200	0.15	5160	0.16	-0.01	0.32
Rural	3200	0.31	5160	0.34	-0.03	0.02
Title I	3110	0.62	4900	0.60	0.02	0.07
District characteristics						

Number of schools	3110	36.90	5110	33.82	3.08	0.15

Appendix Table 3: Balance test results

Variable	Working conditions and job stress			Turnover (binary)			Turnover (three categories)		
	No sanction in 2006-07			No sanction in 2006-07			No sanction in 2006-07		
	Sanction in 2007-08	No sanction in 2007-08	Standardized bias (%)	Sanction in 2007-08	No sanction in 2007-08	Standardized bias (%)	Sanction in 2007-08	No sanction in 2007-08	Standardized bias (%)
Principal characteristics									
Age	49.7	49.5	2.30	49.7	50.7	-11.5	49.7	49.8	-1.00
Female	0.39	0.35	8.80	0.39	0.41	-4.10	0.39	0.35	9.30
Nonwhite	0.17	0.18	-1.80	0.16	0.16	-1.90	0.16	0.15	0.40
Years of principal experience	7.30	7.06	3.50	7.35	7.81	-6.70	7.36	7.60	-3.50
Years in current school as principal	3.82	4.06	-5.10	3.86	3.84	0.50	3.91	4.09	-4.00
Education specialist/doctoral degree	0.41	0.38	6.70	0.41	0.41	0.30	0.42	0.39	5.20
School characteristics									
Distance to the AYP threshold	-25.86	-27.44	7.90	-25.72	-27.26	7.70	-25.87	-27.97	10.50
Percent of nonwhite students	36.38	34.43	6.80	35.55	35.06	1.70	35.10	33.85	4.40
Percent of FRL students	41.81	42.10	-1.20	41.68	43.51	-7.60	41.38	43.06	-7.00
Elementary school	0.25	0.25	0.50	0.26	0.29	-8.00	0.25	0.28	-6.10
Middle school	0.30	0.30	0.00	0.30	0.26	9.60	0.31	0.28	5.50
High school	0.39	0.39	-0.80	0.38	0.39	-2.30	0.39	0.38	1.20
Combination school	0.05	0.05	0.60	0.05	0.05	4.00	0.06	0.06	0.00
School enrollment size	908	840	11.60	904	855	8.40	906	819	14.80
Urban	0.25	0.24	3.10	0.25	0.25	-1.30	0.24	0.22	6.00
Suburban	0.28	0.28	0.00	0.28	0.27	3.00	0.28	0.26	5.10
Town	0.21	0.20	2.30	0.20	0.19	4.10	0.20	0.22	-4.10
Rural	0.27	0.29	-4.70	0.27	0.29	-5.20	0.27	0.30	-6.70

Title I	0.47	0.46	0.50	0.47	0.50	-5.90	0.47	0.48	-3.80
District characteristics								
Number of schools	NA	NA	NA	40.39	43.72	-4.30	40.51	40.97	-0.60

Notes: All variables are based on 2006-07. The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. A total of 45 states are included. Oklahoma has assessment data, but because the scale of their AMO is completely different from that of the other states, I dropped the state from the sample. Standardized biases were calculated using a formula by Rosenbaum and Rubin (1985). NA means that the variable was not used in the first stage model.

Appendix Table 4: Balance test results – combination schools excluded

| Variable | Working conditions and job stress | | | Turnover (binary) | | | Turnover (three categories) | | |
| | No sanction in 2006-07 | | | No sanction in 2006-07 | | | No sanction in 2006-07 | | |
	Sanction in 2007-08	No sanction in 2007-08	Standardized bias (%)	Sanction in 2007-08	No sanction in 2007-08	Standardized bias (%)	Sanction in 2007-08	No sanction in 2007-08	Standardized bias (%)
Principal characteristics									
Age	49.6	50.0	-5.20	49.7	49.6	1.40	49.7	49.6	0.80
Female	0.39	0.36	6.00	0.39	0.37	2.60	0.39	0.37	3.50
Nonwhite	0.17	0.17	-0.40	0.16	0.15	4.10	0.16	0.16	-0.40
Years of principal experience	7.27	7.93	-9.50	7.36	7.39	-0.40	7.34	7.39	-0.80
Years in current school as principal	3.79	4.08	-6.30	3.85	3.83	0.50	3.88	3.84	0.80
Education specialist / Doctoral degree	0.42	0.40	3.50	0.42	0.38	6.60	0.42	0.38	9.30
School characteristics									
Distance to the AYP threshold	-25.86	-27.97	10.50	-25.52	-27.21	8.40	-25.72	-25.95	1.20
Percent of nonwhite students	36.20	33.89	8.10	35.12	33.97	4.10	34.73	35.01	-1.00
Percent of FRL students	41.28	41.15	0.50	41.09	41.28	-0.80	40.74	41.05	-1.30
Elementary school	0.27	0.27	-0.90	0.27	0.30	-5.50	0.27	0.26	0.90
Middle school	0.32	0.33	-1.90	0.32	0.28	9.40	0.32	0.33	-1.70
High school	0.41	0.40	2.60	0.40	0.42	-3.00	0.41	0.41	0.60
School enrollment size	916	829	14.90	907	858	8.40	914	885	5.10
Urban	0.26	0.22	11.10	0.25	0.23	6.90	0.25	0.26	-1.40
Suburban	0.28	0.28	-0.90	0.28	0.28	0.00	0.28	0.26	5.70
Town	0.20	0.21	-2.00	0.20	0.23	-7.60	0.20	0.20	-1.50
Rural	0.26	0.29	-7.30	0.26	0.26	0.30	0.26	0.28	-3.10
Title I	0.45	0.44	1.90	0.46	0.46	-0.30	0.45	0.46	-2.00

District characteristics

Number of schools	NA	NA	NA	39.38	36.08	4.30	39.53	39.57	0.00

Notes: All variables are based on 2006-07. The sample excludes principals working at schools serving grades across school levels (i.e., combination schools). The sample includes only principals who stayed in the same school in 2006-07 and 2007-08 and did not face any sanction in 2006-07. A total of 45 states are included. Oklahoma has assessment data, but because the scale of their AMO is completely different from that of the other states, I dropped the state from the sample. Standardized biases were calculated using a formula by Rosenbaum and Rubin (1985). NA means that the variable was not used in the first stage model.

CHAPTER 5

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

Introduction

Many of the current education policies and programs include some form of accountability incentives such as performance-based sanctions or rewards. Policymakers use these incentives to motivate educators and school administrators to change their behaviors and daily practices in order to achieve their goals including improving student performance and narrowing achievement gaps. Use of accountability incentives will continue to be one of the major education policy options among policymakers at different levels of government. Given this policy trend, it is pivotal to understand how educators and school administrators respond to these incentives. Without such knowledge, policymakers may design an incentive policy or program inappropriately, and therefore encounter many unintended consequences and eventually hurt students' chances of success in their future. Although a large number of studies have examined the effect of NCLB on students and teachers (e.g., Ballou & Springer, 2011; Dee & Jacob, 2011; Grissom, Nicholson-Crotty, & Harrington, 2014; Lauen & Gaddis, 2012; Reback, Rockoff, & Schwartz, 2014), fewer studies have investigated the effect on the school principals. This dissertation contributes to our understanding of how school principals respond to accountability incentives. It focuses on the principals' working conditions, job stress, and turnover behaviors. It provides some of the first empirical evidence on principal behaviors under NCLB sanctions.

In this chapter, I first review the findings of this research. Following the review, I discuss the implications of my findings for policymakers. I conclude with directions for future research.

Study 1: Influence of NCLB: A Review of Literature, Theories behind the NCLB Accountability System, and An Examination of the System

The NCLB accountability system was implemented in 2002 under the assumption that its sanction system would motivate educators and administrators to change their behaviors and daily practices so that policymakers can achieve their goals such as improving the overall student performance and narrowing achievement gaps among different student subgroups. This study reviews prior studies on the effect of NCLB and examines the design of the NCLB accountability system using theories of motivations and incentives.

The literature review generally finds that NCLB improved student performance, especially in mathematics and among minority and disadvantaged students (e.g., Dee & Jacob, 2011; Lauen & Gaddis, 2012; Wong, Cook, & Steiner, 2009). Its effect on teachers appears to be mixed. For example, the review finds that under NCLB teachers had high academic expectations for student learning, increased their effort to help students learn, especially low-performing students, and worked for longer hours (Finnigan & Gross, 2007; Guggino & Brint, 2010; Mintrop, 2004; Murnane & Papay, 2010; Reback et al., 2014). On the other hand, the review also finds that teachers had unfavorable views toward NCLB and that NCLB gave teachers an increasing amount of pressure and caused demoralization among them (e.g., Berryhill, Linney, & Fromewick, 2009; Center on Education Policy, 2006; Dee, Jacob, & Schwartz, 2013; Educational Testing Service, 2008; Guggino & Brint, 2010; Mertler, 2011; Santoro, 2011; Sunderman, Tracey, Kim, & Orfield, 2004). Teachers felt less secure about their job and some of

them decided to leave their schools (Deniston & Gerrity, 2010; Feng, Figlio, & Sass, 2010; Reback et al., 2014; Santoro, 2011; Sunderman et al., 2004).

The review reveals a lack of studies about the effect of NCLB on school principals. Most of the extant studies use qualitative data and focus on small samples of school principals. These studies report that NCLB made principalship more challenging and that principals viewed the NCLB system as an unfair system (e.g., Brown, 2006; Lyons & Algozzine, 2006). The review suggests that NCLB generally impacted school principals negatively. Yet, fewer studies have systematically examined its effect on principals.

While a large number of studies have investigated the influence of NCLB, few studies examine the design of the NCLB accountability system. Using theories of motivations and incentives, I assessed whether the system was appropriately designed and identified problems in the system. First, performance measures used in the system do not appear to have reflected the true values of job performance among educators and administrators. Because the teaching profession and the principalship have multi-tasking features (Podgursky & Springer, 2007), the percent of proficiency achieved only captures one dimension of their jobs. Second, the performance measures and criteria used to determine a sanction status might have conflicted with the educators' professional standards and personal educational philosophies, as survey studies report that teachers and principals had unfavorable views toward the NCLB system and perceived it as an unfair system (Educational Testing Service, 2008; Guggino & Brint, 2010). Third, the performance measures used in the system emphasized quantity, not quality. This emphasis might give educators less autonomy over how to educate their students. Fourth, the system applied the same performance targets to all schools, regardless of differences in school contexts, although the Safe Harbor considers growth trajectories in school performance. Finally,

228

the first two years of the NCLB sanctions might not have given adequate incentives for educators and principals to respond because during the first two years districts offered technical assistance and supplemental educational services (SES) to failing schools. District technical assistance and SES could have been perceived as rewards rather than sanctions. All of these problems could contribute to the inefficiency or ineffectiveness of the system. These problems could prevent the system from adequately motivating educators and principals to change their behaviors.

This analysis highlights the importance of the design in an incentive policy. If the policy is inappropriately designed, it is likely to backfire and encounter unintended consequences. Policymakers need to make decisions about the incentive design carefully. For example, they would need to understand the nature of tasks that educators and administrators perform to determine what performance measures to use. They would also need to consider whether use of a single performance target is appropriate to incentivize all school staff because school contexts are substantially different from school to school. One performance target could be relatively easy for certain types of schools; it may be more challenging for other types of schools. Differentiated performance targets may be more effective under these circumstances. Overall, incentives work only when policymakers craft the design carefully.

Study 2: Principal Turnover under NCLB Accountability Pressure

The first study identified a lack of studies about the influence of NCLB on school principals. Little is known about how school principals respond to NCLB sanctions. For example, we do not know much about whether school principals systematically leave their schools when they face the sanctions. Neither do we know whether they choose to transfer to different schools, change their positions, or exit the public education system. Along with this inquiry, it is also

unclear whether principals systematically transfer to certain types of schools or whether they systematically take up certain jobs. Moreover, there is ambiguity in understanding whether principals' turnover behaviors differ by their qualifications and characteristics of the schools in which they work.

This empirical study contributes to the current literature by providing some of the first empirical evidence on these research questions and informs our understanding of principals' turnover behaviors under NCLB sanctions. I have used longitudinal administrative data and detailed school-level assessment and AYP data obtained from Missouri Department of Elementary and Secondary Education. A variable that measures the distance to the AYP threshold was constructed in this study and was used as a key matching variable along with other principal, school, and district characteristics in order to identify statistically comparable school principals except their schools' sanction status.

I find that the informal sanction is not associated with higher turnover rates, irrespective of facing the informal sanction for the first time or not. Yet, principals' responses to the informal sanction are different based on their job experience level and the percent of nonwhite students at schools they serve. For example, I find that inexperienced principals are more likely to turn over when they face the informal sanction, including both first time and non-first time, than their colleagues with the same experience level but not facing it (24% and 11%, respectively). This difference becomes smaller or even turns into negative, as their experience level increases. Similarly, I find that the difference in the predicted probability of turnover between principals facing the informal sanction and those not facing it becomes larger as the percent of nonwhite students increases. The difference is minus two percentage points among principals serving the smallest number of nonwhite students; it jumps to 17 percentage points among principals serving

230

the largest number of nonwhite students (75% to 100%). These moderation effects are also observed by turnover types. For example, inexperienced principals are more likely to transfer to different schools when they face the informal sanction. Similarly, they are more likely to change their positions.

In contrast, I do not find evidence that the formal NCLB sanctions (i.e., School Improvement Year 1 and beyond) are associated with principal turnover. Neither do I observe any moderation effects by principal qualifications nor school characteristics. Moreover, I find no evidence that principals' transfer patterns or position-change patterns are systematically different by the sanction status. All principals who choose to transfer or change their positions behave similarly, whether they face the informal sanction or not.

As discussed throughout this dissertation, the extant qualitative studies suggest that NCLB generally impacted principals negatively. Furthermore, two recent empirical studies find that principals are more likely to turn over when they face NCLB sanctions (Ahn & Vigdor, 2014; Li, 2012). Results from my analysis expand these findings by providing new evidence that principals, especially inexperience principals and principals serving a large number of nonwhite students, are more likely to leave their schools, transfer to different schools, or change their positions when they face the informal sanction.

However, my analysis does not find evidence that the formal NCLB sanctions influenced principals' turnover behaviors. This finding could be explained by the design of the sanction system. As discussed in the first study, one of the problems in the NCLB sanction system is that the first two years of the sanctions might not have provided adequate incentives for principals to change their behaviors, or turnover behaviors in this context, because the sanctions include district technical assistance and supplemental educational services. Principals might have

231

perceived these sanctions as rewards, rather than sanctions. If this was true, then these sanctions might have motivated principals to *stay* rather than leave their schools. It is critical to understand how principals perceived these sanctions and the NCLB sanction system.

Study 3: Principals' Working Conditions, Job Stress, and Turnover Behaviors under NCLB Accountability Pressure

While the second study focuses on principals and their turnover behaviors in the state of Missouri, this study expands the sample of principals to a national representative sample of principals and examines whether NCLB sanctions influenced principals' working conditions, job stress, and turnover behaviors. I focus on the informal sanction and the first year of the sanction or School Improvement Year 1 (SIY 1) and combine them together. I use data from the Schools and Staffing Survey 2007–08 and merge them with detailed school-level assessment and AYP data systematically collected from 45 states.

First, for principals' working conditions and job stress, I find evidence that the NCLB sanction is associated with higher job stress but no evidence that it is correlated with the principals' working conditions. Job stress is especially greater among principals serving a large number of nonwhite students and among inexperienced principals. Second, for principals' turnover behaviors, I find that principals are more likely to leave their schools or transfer to different schools when they face the sanction. For example, the predicted probability of principal turnover is 23 percent when a principal faces the sanction; it is only 16 percent when the principal does not face it. However, unlike the second study in Missouri, I find little evidence that these associations are moderated by principal qualifications or school characteristics. All of these patterns still hold after accounting for the influence of state sanction systems.

Given these findings that the NCLB sanction is associated with higher job stress and turnover rates, I examine whether job stress functions as a mediator between the sanction and principal turnover. My causal mediator analysis finds little evidence that it mediates the relationship. This finding is somewhat surprising and contrary to what the literature suggests. The significant relationship between the sanction and turnover largely comes from its direct effect. Yet, it is still possible that there are unobserved mediators that connect the NCLB sanction and principal turnover. Moreover, the null result also suggests that turnover could be driven by district administrators.

Not all of these findings are consistent with those found in the second study. For example, in Missouri, I find that the informal sanction is in general not associated with higher turnover rates. However, the relationship is moderated by years of principal experience and the percent of nonwhite students. I find that inexperienced principals are more likely to leave their schools when they face the informal sanction, whether it is the first time or not. Similarly, principals serving a large number of nonwhite students are more likely to turn over when they face the informal sanction. These patterns are also found in transfer rates and position change. Yet, the formal NCLB sanctions are found uncorrelated with any turnover behaviors.

At the national level, I find that principals are more likely to transfer to different schools when they face the NCLB sanction including both the informal sanction and School Improvement Year 1. I find little evidence that this relationship is moderated by principal qualifications or student demographics.

These differences could be due to a difference in what sanction is considered. While I examine the informal sanction and the formal NCLB sanctions separately in the second study, I combine the informal sanction and School Improvement Year 1 (SIY 1) in the third study. The

233

inclusion of SIY 1 might have caused these differences. Moreover, Missouri is a relatively rural state, with 40 percent of schools located in rural areas, compared with 34 percent in a median state. Although the state is a national representative among other important school characteristics, the rural context might have influenced the principals' turnover behaviors.

Summary of Contributions

In short, my dissertation contributes to our understanding of how school principals respond to accountability pressure. First, I offer a framework to assess the design of incentive policies and programs in general and identify problems in the design of the NCLB accountability system based on theories of motivations and incentives. These problems are also often found in other incentive policies and programs in education. The framework can be used for policymakers to design an appropriate incentive policy and improve the effectiveness of the policy. Second, I provide empirical evidence about the relationship between NCLB sanctions and principal turnover and offer detailed analyses of principal turnover behaviors. The extant literature only examines whether a principal turns over when he or she faces an NCLB sanction. The second study digs deep into the relationship and investigates each NCLB sanction including the informal sanction, different turnover types, moderation effects by principal qualifications and school characteristics, and transfer and position-change behaviors. My findings suggest that inexperienced principals and those serving a large number of nonwhite students are more likely to turn over, transfer to different schools, and change their jobs. Finally, I examine the influence of NCLB sanctions on principals at the national level. This study provides what is, to my knowledge, the first empirical national representative evidence of the relationship between NCLB sanctions and principals' working conditions, job stress, and turnover behaviors. While

the extant literature provides some anecdotal evidence about the relationship, it is mostly based on small samples of principals, survey studies, opinion essays, and/or a single state. My study is the first to use a nationally representative sample of school principals to examine the relationship. I find that principals feel greater job stress when they face the NCLB sanction. I also find that principals tend to transfer to different schools when facing the sanction, although I do not observe moderation effects by principal qualifications or school characteristics.

Implications for Policymakers

School leadership is "the second only to classroom instruction among school-related factors that affect student learning in school" (Wallace Foundation, 2013, p. 5). They play a critical role in improving school and student outcomes (e.g., Branch, Hanushek, & Rivkin, 2012; Brewer, 1993; Grissom, Kalogrides, & Loeb, 2015; Grissom, Loeb, & Master, 2013; Hallinger & Heck, 1998; Waters, Marzano, & McNulty, 2003). It is, therefore, important for policymakers to understand how they respond to accountability pressure and, in more general, incentives, to design an incentive policy. Findings from this dissertation provide policymakers with some useful information for this purpose.

First, policymakers should carefully craft the design of incentives. For instance, they need to understand the nature of tasks that school principals work on. Depending on whether the tasks are multi-dimensional, policymakers should decide whether they use a single performance measure or multiple quantitative and qualitative performance measures. They also need to consider performance targets carefully. If all principals are currently performing at the same level, policymakers can use a single performance target. However, if their performance varies from principal to principal and from school to school, they have to consider differentiated

235

performance targets. A single performance target would not motivate principals currently far above or below the target. Moreover, policymakers need to understand whether sanctions or rewards that they choose to use provide adequate incentives for principals to respond. For instance, if the size of incentives is not significant, principals would not respond.

Second, my findings suggest that principals tend to feel greater job stress when they face accountability pressure, especially when they do not have enough job experience or when they serve a large number of nonwhite students. Since the literature suggests that job stress leads to burn out, which then leads to turnover (Leiter & Maslach, 2009), policymakers need to provide professional support for principals facing the pressure, especially for inexperienced principals and principals serving a large number of underprivileged students. Professional support could include induction/mentoring programs, assignment of additional school support staff such as school improvement coaches and curriculum experts, and professional development programs.

Third, my analysis suggests that principals tend to turn over, transfer to different schools, and change their positions when they face accountability pressure. These patterns are observed among inexperienced principals and those serving a large number of nonwhite students. Principal turnover is not necessarily a bad thing for schools if ineffective principals leave or district administrators strategically reallocate principals to different schools. Yet, if effective principals voluntarily leave their schools when they face the sanction, this becomes a challenge for school improvement efforts. Again, policymakers would need to provide some form of incentives to retain them. They could offer a retention bonus or change principals' working environments by providing additional funding and/or school staff to lighten the principals' burden.

Overall, policymakers always need to pay careful attention to the design of an incentive policy. If it is not appropriately designed, principals, teachers, and eventually students will suffer

from it. To avoid unintended consequences, policymakers would need to communicate with school administrators and educators in advance and conduct a pilot study in order to decide the design of the policy.

Directions for Future Research

This dissertation suggests future research projects. First, few studies have assessed the design of an accountability system. Although my dissertation focuses on NCLB, future research should also analyze the design of incentives embedded in state and local accountability systems. Accountability systems are now widely used at different levels of government and affect many educators, administrators, students, and parents. Given its influence, this research might benefit all of the stakeholders. A case study approach may be particularly well suited to assess the design.

Second, my findings about the relationship between the NCLB sanction and principals' job stress suggest that accountability pressure may also influence the relationship between principals and teachers because the current school reform emphasizes the importance of instructional leadership by principals. Principals are expected to work closely with teachers to improve student performance. Under this leadership environment, if principals feel greater job stress due to accountability pressure, it is likely that their interactions with teachers are also influenced. Findings from this research may provide some insights about how to build a good instructional relationship with teachers.

Third, my dissertation focuses on the short term influence of NCLB or accountability pressure. However, its influence could be cumulative. If a principal faces accountability pressure longer, his or her behaviors may be quite different from their short-term behaviors. While

longitudinal data on principals' behaviors may not be available at the national level, they may be available at the state level. Results from this research will help district administrators to make strategic human resource decisions. For example, they may regularly reallocate principals facing accountability pressure for long periods to high-performing schools to reduce their job stress and retain them in the district.

Finally, this line of research as well as my dissertation can be applicable to teachers and district administrators (e.g., district superintendents). Although many more researchers have investigated the influence of NCLB on teachers than school principals, few studies have empirically examined the relationship (e.g., Grissom et al., 2014). At the district level, to my knowledge, almost no studies have empirically investigated the influence of NCLB, or more generally, accountability pressure. Accountability systems, whether at the federal or state level, target not only educators and school administrators but also district administrators. Knowing how district administrators, especially district superintendents, respond to accountability pressure would also benefit policymakers in designing the systems.

References

Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina*. NBER Working Paper No. 20511. Retrieved in October, 2014, from http://www.nber.org/papers/w20511. Cambridge, MA: National Bureau of Economic Research.

Ballou, D., & Springer, M. G. (2011). *Has NCLB encouraged educational triage? Accountability and the distribution of achievement gains*. Unpublished manuscript. Nashville, TN: Vanderbilt University.

Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The Effects of Education Accountability on Teachers: Are Policies Too-Stress Provoking for Their Own Good? *International Journal of Education Policy and Leadership, 4*(5), 1-13.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals. NBER Working paper No. 17803. Retrieved from http://www.nber.org/papers/w17803. Cambridge, MA: National Bureau of Economic Research.

Brewer, D. J. (1993). Principals and student outcomes: Evidence from U.S. high schools. *Economics of Education Review*, *12*(4), 281-292.

Brown, P. F. (2006). Preparing principals for today's demands. *Phi Delta Kappan*, *87*(7), 525-526.

Center on Education Policy. (2006). *Ten big effects of the No Child Left Behind Act on public schools*. Retrieved from www.cep-dc.org. Washington, DC: Center on Education Policy.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis, 35*(2), 252-279.

Deniston, R. D., & Gerrity, K. W. (2010). Elementary school teachers' perceptions of No Child Left Behind and its effect on morale. *scholarlypartnershipedu*, *5*(2), article 4.

Educational Testing Service (2008). *Standards, accountability, and flexibility: Americans speak on No Child Left Behind.* Princeton, NJ: Author.

Feng, L., Figlio, D., & Sass, T. (2013). *School accountability and teacher mobility*. Working paper. Atlanta, GA: Georgia State University.

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

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3-28.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433-444.

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers and their work environment. *Educational Evaluation and Policy Analysis, 36*(4), 417-436.

Guggino, P. C., & Brint, S. (2010). Does the No Child Behind Act help or hinder K-12 education? *Policy Matters*, *3*(3), 1-7.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*(2), 157-191.

Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis*, *34*(2), 185-208.

Leiter, M. P., & Maslach, C. (2009). Nurse turnover: the mediating role of burnout. *Journal of Nursing Management*, *17*(3), 331-339.

Li, D. (2012). *School accountability and principal mobility: How No Child Left Behind affects the allocation of school leaders*. Working paper. Cambridge, MA: Harvard University.

Lyons, J. E., & Algozzine, B. (2006). Perceptions of the impact of accountability on the role of principals. *Education Policy Analysis Archives*, *14*(16), 1-16.

Mertler, C. (2011). Teachers' perceptions of the influence of *No Child Left Behind* On Classroom Practices. *Current Issues in Education*, *13*(3). Retrieved on July 20, 2014, from http://cie.asu.edu/

Mintrop, H. (2004). *Schools on probation: How accountability works (and does not work)*. New York, NY: Teachers College Press.

Murnane, R. J., & Papay, J. P. (2010). Teachers' views on No Child Left Behind: Support for the principles, concerns about the practices. *Journal of Economic Perspectives*, *24*(3), 151-166.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, *26*(4), 909-949.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, *6*(3), 207-241.

Santoro, D. A. (2011). Good teaching in difficult times: Demoralization in the pursuit of good work. *American Journal of Education*, *118*(1), 1-23.

Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom realities and No Child Left Behind*. The Civil Rights Project. Cambridge, MA: Harvard University.

Wallace Foundation. (2013). *The school principal as leader: Guiding schools to better teaching and learning*. Washington, DC: Author.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced Leadership: What 30 years of research tells us about the effect of leadership on student achievement* (Working paper). Aurora, CO: Mid-continent Research for Education and Learning.

Wong, M., Cook, T. D., & Steiner, P. M. (2009). *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series*. Northwestern University Institute for Policy Research Working Paper. Evanston, IL: Northwestern University.