An Investigation into the Impact of the Differential Implementation

of Accountability Policy on the Teacher Experience

_____

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

David N. Martin, BS, MS

May, 2020

Abstract

This dissertation investigates the impact of variations in accountability policy implementation on teachers. Past studies (Dee & Jacob, 2010, 2011) have investigated how the introduction of a new accountability policy impacts teachers, but they do not take into account the ongoing and extensive variations in how states implemented their individual policies. My paper examines how these variations, as quantified by a unique stringency measure created by Wong, Wing, Martin, and Krishnamachari (2017), impact the teacher experience during No Child Left Behind (NCLB). Utilizing this stringency measure, I estimate the causal effects of the changing and ever increasing difficulty of accountability policies on teacher labor force outcomes, perceptions, and attitudes. I also look at how varying levels of teacher experience, school urbanicity, and other characteristics moderate the effect of varying stringency levels. Overall, the results do not suggest an overwhelming impact of increasing stringency on teachers, however, future teachers' decisions about going into teaching had a strong negative association with increasing stringency. The analysis also shows that stringency differentially impacted teacher outcomes based upon teacher and school characteristics, as well as, past accountability policies at the state level.

Education Leadership, Foundations and Policy

Curry School of Education

University of Virginia

Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, ("An Investigation into the Impact of the Differential Implementation of

Accountability Policy on the Teacher Experience"), has been approved by the Graduate Faculty

of the Curry School of Education in partial fulfillment of the requirements for the degree of

Doctor of Philosophy.

_____

Name of Chair (Vivian Wong)

_____

Committee Member Name (Daphna Bassok)

_____

Committee Member Name (Dan Player)

_____

Committee Member Name (Peter Youngs)

_____Date

DEDICATION

This work is dedicated to my wife for the many years she has supported me on this journey and

to my daughter Callie for providing plenty of laughs along the way.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee for their patience and valuable insight during this dissertation process. I especially want to thank to Dr. Wong for her long-term commitment to me and to this project. Her input and words of wisdom have helped to provide me with the fortitude I needed to keep moving forward.

I also want to thank the long line of fellow students, researchers, and co-workers who have helped me along the way. To Theresa and Kate, thank you for always providing an ear, allowing me to express my reservations, confusions, and thoughts about this process. I would also like to thank Justin and Anandita for their assistance in getting the project to this point and for continuing on with the work. Finally, I would like to thank my coworkers at the Health Sciences Library for their input and for their patience while I worked towards the completion of this dissertation.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

Over the last 20 years, school accountability policies have been the centerpiece of educational reform efforts in the United States. These policies hold schools directly "accountable" for student performance. The rationale here is that by aligning achievement standards with testing and accountability measures, schools have strong incentives to improve student performance. Most accountability measures involve annual reporting of aggregate student achievement at both the school and district levels, and some form of remediation for schools and districts that miss targets in proficiency subjects. Corrective actions may include increased oversight by state departments of education, additional resources for teacher professional development and student tutoring services, and possible reconstitution or closure of schools that persistently fail to meet academic performance standards.

Critics of accountability policies argue that these reform efforts make it more difficult for schools to attract and retain high quality teachers, which may negatively affect student achievement and learning. Proponents of the policies, however, argue that schools and school districts that are well aligned with higher performance standards will attract good teachers who will want to stay in their jobs, improving student outcomes in the longer term. The most consequential of the reform initiatives was No Child Left Behind (NCLB). Introduced in 2001 and fully enacted in 2003, NCLB centered on a push to ensure that all students in grades 3-8 were 100% proficient in reading and math by 2014 based on performance on state standardized tests. Failure to reach the 100% proficiency goal and other intermediate objectives would have potentially resulted in sanctions for states, school districts and schools, putting additional pressures on all stakeholders.

However, states had considerable leeway in implementing policies to help schools achieve the 100% proficiency goal. For example, states were required to set annual targets (in percentage of students proficient) to help schools and districts meet the federal mandate, but they had discretion over which measures that were used to assess students' proficiency, the trajectory of improvement that schools were required to follow, and so-called "exemption" rules that may be applied to schools and districts, which effectively lowered the proficiency requirements for many schools. Taken together, these state-level decisions introduced tremendous heterogeneity in the amount of pressure that each state places on schools to improve.

Until recently, research on accountability policies (Dee & Jacob, 2011) has focused on the impact of introducing NCLB on school, teacher, and student outcomes. The headline result from these studies suggests that NCLB had substantial impacts on students' math achievement but not on reading, on teacher compensation, and on the percentage of teachers with graduate degrees (Dee & Jacob, 2010, 2011; Wong, Steiner, & Cook, 2015). However, NCLB did not end after the initial introduction of accountability reform. In fact, NCLB was in place for more than a decade until it was eventually replaced by the Every Students Succeed Act (ESSA) in 2015, another consequential accountability policy. As mentioned, much of the previous research focused on consequential accountability policies, such as NCLB or ESSA, as a shock to the system. This shock may not capture the full impact of accountability and it may be possible that any impacts initially felt by a new policy are short lived. Mark Schneider (2011) introduces the idea of an "Implementation Plateau" where early gains are found from the implementation of a new policy, but these gains start to stagnate, and often, turn to losses. Implementation Plateaus have been studied in student testing and scores, but perhaps it applies to teachers as well. It is unclear whether teachers continue to benefit from mounting accountability pressures under

NCLB and this study will provide a way to test whether teachers stop responding to increasing stringency. This dissertation addresses this issue by utilizing a novel measure of state accountability stringency measuring policy implementation between the years of 2003 and 2011 to assess the impact of increasing accountability pressures on teacher employment and job satisfaction outcomes.

Teachers' Responses to Accountability Pressures

During NCLB, much of the heavy lifting for improving student performance in reading and math fell to teachers. Without teachers who can facilitate the learning process, any policy implementation strategies would be for naught. Teachers provided the most direct impact on student scores and were held responsible if students did not show adequate gains towards full proficiency. Therefore, it is critical to understand how teachers responded to increasing accountability pressures under NCLB. Prior research found that increased responsibility over student performance led to stress for many teachers and to efforts to improve student performance through various means, both ethical and unethical, legal and illegal (see, for example, Dee, Jacob, & Schwartz, 2013; Grissom, Nicholson-Crotty, & Harrington, 2014; Jacob, 2005). For example, teachers changed pedagogical practices in order to align their teaching content with what was being taught on the standardized assessments (Dee, Jacob, & Schwartz, 2013; Reback, 2008; Reback, Rockoff, & Schwartz, 2014). Many teachers who did not directly teach math or reading were expected to teach these subjects during their own class periods (Booher-Jennings, 2005; Figlio & Rouse, 2006). Teachers also discovered ways to identify students who were in danger of failing, and delivered more specific instruction to those students in order to push them just past the proficiency threshold (Neal & Schanzenbach, 2010). The above responses illustrate the notion that teachers did indeed feel an impact from NCLB;

however, it is unclear how directly NCLB impacted teachers outside of these pedagogical responses. More specifically, how did mounting accountability pressures under NCLB impact teachers' employment and job satisfaction outcomes?

Accountability Stringency and Teacher Workforce Outcomes

This study utilizes a novel measure of state accountability stringency introduced by Wong, Wing, Martin, and Krishnamachari (2017) that describes the complicated array of ways in which states implemented NCLB. To develop this measure, a database of state accountability rules from 2003 to 2011 (i.e., the NCLB pre-waiver period) was created. The database was then used to develop a proficiency calculator that would determine how a particular school would be evaluated under the rules in each state and year. Employing the proficiency calculator, Wong, et al (2017) calculated the percentage of a fixed sample of schools that would have failed to meet the standards for Adequate Yearly Progress (AYP) in each state and year. The simulated failure rates in the fixed sample provide a concrete measure of the stringency of each state's NCLB implementation. The stringency measure takes account of most AYP decisions made by the state, but is independent of school attributes and teacher and student characteristics in the state. Creating a measure that is independent of school, teacher, and student characteristics is important because it separates the issue of implementation stringency from actual teacher outcomes.

In this dissertation, I use this implementation measure to describe state accountability stringency under NCLB, and to examine the causal impact of variations in accountability stringency on teacher employment outcomes. In particular, I explore how teachers move in and out of the profession, their perceptions of their job, and their attitudes towards being a teacher as accountability pressures mount under NCLB. Moreover, I examine whether fewer prospective teachers enter the profession as accountability stringency is increased across the states. Research

shows that teacher turnover and teacher labor shortages greatly impact both student outcomes and the student experience (Clotfelter, Ladd, & Vigdor, 2011; Ronfeldt, Loeb, & Wyckoff, 2013), so it is essential to understand the relationship between accountability stringency and the teacher experience.

Conceptual Model of How NCLB Affects Teachers

How does increasing state accountability stringency affect teacher outcomes? There are two distinct paths that describe teachers' responses to increased accountability pressures. On one path, the negative path, stringency leads to a decline in the quality of working conditions through higher demands being placed on teachers by means of more paperwork and more hours of planning and lower support from administrators and fellow teachers. This leads to teachers feeling more stressed and less control over what they teach and within the classroom. This encourages teachers to leave the teaching profession altogether or to move to a new school or state. This inevitably can lead to the teaching profession as a whole being seen as a less desirable occupation to possible new teachers. This path takes a negative point of view on NCLB and if this is the predominant reaction by teachers, it would illustrate how increasing NCLB stringency would lead to an overall poorer experience of being a teacher.

Conversely, NCLB stringency may be linked to a more positive impact on teachers. Changes in NCLB stringency may lead to an improvement in working conditions by way of a better alignment to goals and to higher support from colleagues and administration. The daily experience of being a teacher may also be improved through monetary incentives, more autonomy and control within the classroom, and recognition for student test score improvements. This in turn would encourage less teacher turnover and, higher retention of teachers within the

profession and would lead to the increased attraction of teaching as an occupation, leading to greater numbers of new teachers.

Research on the Impact of Accountability Policy on Teachers

Prior research on the impact of accountability policy and on perceived responses by teachers to accountability policies comes with issues of generalizability and little empirical evidence about how increasing accountability pressures affected teacher outcomes. One common issue in much of this research is the use of a single state or district to investigate accountability policy and responses to accountability policy. This research provides a glimpse into what occurred within a single location or a small set of locations (Feng, Figlio, & Sass, 2010; Neal & Schanzenbach, 2010), however, it does not provide a wider glimpse across a swath of locations and the research may only have applicability and validity to these specific locations. For example, Feng, Figlio, and Sass (2010) focused solely on the impact of consequential accountability and increased or decreased accountability pressure on teachers within the state of Florida. Another common theme in the research is the focus on the increased pressure on teachers when students were at or near a specific proficiency threshold (Neal & Schanzenbach, 2010; Reback, Rockoff, & Schwartz, 2014). Neal and Schanzenbach (2010) studied those students who fell right near the pass or fail threshold and found that teachers were focusing more on these students were applying more focus towards these students and changing their teaching approaches.

One other common thread through much of the accountability policy research is the use of analytical methods that look at the presence or absence of accountability policies before the implementation of No Child Left Behind in 2001 (e.g., Dee & Jacob, 2011; Grissom, Crotty, & Harrington, 2014). Dee and Jacob (2010, 2011) examined the variable impact of the introduction

of NCLB on student achievement and teachers within states with and without accountability policies prior to 2001. They found teacher compensation, teachers with master's degrees and the amount of time spent on math and English all significantly increased in states without prior consequential accountability policies compared to those with consequential accountability policies. These studies utilized a binary indicator for whether or not a previous accountability policy was in place, which did not provide a granular measure of the stringency of NCLB in each state and year. The introduction of a new accountability policy is definitely going to have a large impact on teachers, however, these impacts may not last (Schneider, 2011). The goal of this dissertation is to fill the gap in the research and investigate the effect of variable implementation stringency and the total suite of implementation policies enacted within each state and year.

Research Questions

This dissertation address gaps in the existing literature on the impact of changes in accountability stringency on teacher outcomes. To do so, I evaluate the following three questions:

1. What impact did increased accountability stringency have on teacher labor market outcomes?

2. What impact did increased accountability stringency have on teachers' attitudes and perception of their work?

3. Did increased accountability stringency differentially affect teacher outcomes based on school characteristics, such as the urbanicity of the school, and teacher characteristics, such as years of experience and ethnicity?

To address these questions, I take advantage of the extensive variation in state implementation stringency under NCLB. Under this federal policy, a school that failed to reach proficiency in one state using one set of rules may be deemed proficient in another state that utilizes a different set of rules (Wong et al., 2017). I use the previously-discussed measure of stringency that gauges how the same sample of schools would perform against different variations of NCLB implementation and strips away the impact of student characteristics and performance, leaving only the difference in state and year policy implementation.

This dissertation examines how states' accountability stringency impacts teacher outcomes gathered from U.S. Census data, the School and Staffing Survey (SASS), the Teacher Follow-up Survey (TFS), the Freshman Survey, and the College Senior Survey. The data used here provides a holistic snapshot of teacher experiences under NCLB, including the practice of working as a teacher and decisions to enter or leave the workforce. To estimate the impact of stringency variation on teacher outcomes, I utilize a two-way fixed effects approach that helps to control the impacts of state and time on teachers, while isolating the effect of stringency. The fixed effects approach, in conjunction with a measure of stringency, allows me to investigate if teachers in highly-stringent states felt more or less influence of NCLB than those teachers in states with a low level of stringency. The results of this study may provide evidence as to why teachers are entering the profession at a decreased rate and inform policy-makers on how best to retain teachers.

I estimate the effects of variation in NCLB stringency on teacher labor market outcomes, as well as, teacher perceptions and attitudes using a two-way fixed effects approach. The findings suggest that increasing accountability stringency does not have a measurable impact on teacher's choices to leave the teaching profession, but it does have a modest impact on teachers'

choosing to move away from their current school. The analysis shows clear differentiation among subgroups in regards to teacher movement with female teachers showing a higher likelihood of leaving the profession with increasing stringency and middle-aged teachers (ages (36-50) showing the opposite relationship.

As far as perceptions and attitudes are concerned, very little overall impact was found, other than the increase in salary dissatisfaction as stringency increases. These impacts were felt differentially, however, with some subgroups feeling a negative impact of increasing stringency on support from colleagues and administrators (female, older and more experienced teachers). For other subgroups, there was a positive impact of increasing stringency on control in the classroom and administrative support (rural, male, younger teachers). Increasing stringency had a significant effect on future teachers; for college freshmen, increasing stringency was associated with a decrease in desire to pursue an education major or an education career. The analysis also indicates that teachers in states that had accountability policies prior to NCLB were differentially affected by increasing amount of stringency while this federal policy was in place. States that moved from strong or moderate consequential accountability policies to NCLB consistently showed a plateauing or even a negative impact of increasing stringency on several key teacher outcomes.

BACKGROUND

Research has focused on the impact of NCLB and other accountability policies on student performance, school resource changes, and instructional and curricular modifications by schools and districts, as well as the ways that teachers experienced accountability pressures. This chapter examines prior literature on pressures teachers faced under NCLB and their responses. Overall, research suggests that teachers felt pressure under accountability reform that was both positive and negative. Research indicates that accountability pressure led to teachers taking more control within the classroom and receiving more support from administration, however, other research shows that accountability pressures led to teachers teaching to the test, focusing on tested subjects, being disenchanted with the focus on standardized assessments, and changing their practice.

The review proceeds as follows. First, I provide an overview of the accountability literature from the last 30 years, including the introduction of NCLB. Although NCLB is often described as a "one-size-fits-all" approach, states had considerable leeway in holdings schools to different accountability expectations. Second, I describe the multiple implementation options that states had under NCLB. Third, I consider prior research describing teachers' experiences under NCLB, including their perceptions of the teaching profession, control within the classroom, influence over choices within the school, and other important facets of their professional experience. I conclude by summarizing existing gaps in the accountability literature and teacher responses.

Holding Schools Accountable for Student Performance

The objective of accountability policies, at their most basic level, is to increase student achievement by holding schools responsible for providing students with a suitable learning

environment through the means of rewards and sanctions tied to various measures and performance standards (Figlio & Loeb, 2011). Although this basic description fits the accountability policies of the past several decades, the implementation of these accountability policies differs in a number of distinct and diverse ways.

       Pre-NCLB Accountability policies. The movement towards accountability stems from a push in the 1980s by Ronald Reagan's administration to gauge performance in both the public and nonprofit sectors (Figlio & Kenny, 2009). This push filtered into the standards-based reform movement of the time and created a need for accountability within schools, districts, and states (O'Day & Smith, 1993).The reauthorization of the Elementary and Secondary Education Act (ESEA) in 1994, as well as the Improving America's School Act (IASA) and Goals 2000 stem from that push. These policies encouraged states to increase graduation rates, raise student achievement, and improve teacher quality by the year 2000 and Goals 2000 even required states to implement content standards in order to receive Title I funds. These policies recommended rigorous standards and assessments, however they were not mandated, allowing states and localities to reach their goals using a variety of strategies and their own specific policies (Dee & Jacob, 2010). These policies varied greatly between the states, with just over half of the states (26), largely concentrated in the South, utilizing consequential accountability policies or policies that tie sanctions and rewards to student and school performance by 2000 (Hanushek & Raymond, 2005). In the 1990s, only a few states and locales had an accountability or testing policy that applied across grades 3-8, including North Carolina, Texas, and Chicago. This is one of the key reasons much of the research on consequential accountability policy prior to NCLB was done in these specific locations.

While states varied in the types of sanctions that were introduced to schools, some sanctions, although not universally utilized, included the introduction of outside intervention teams to monitor student progress, the provision to allow families to choose alternative school options, and the threat of school reconstitution, take-over, or closure (Insights on Educational Policy, Practice, and Research, 2000). Four states simply used Report Card accountability where states made school level performance public, but did not have the sanction and reward system. Twenty other states did not set up an official accountability policy that provided reports on performance or required sanctions and rewards. Although federal accountability policy during this time provided states with goals and certain guidelines, it did not provide a direct framework or metric by which states should be held accountable.

Introduction of No Child Left Behind. In 2002, No Child Left Behind was passed and signed into law by George W. Bush. The policy was based on Texas' framework of consequential accountability that was in place during Bush's time as governor in Texas. The enactment of NCLB replaced Goals 2000, which had focused on having states enact content standards, and put into place a framework that called for the disaggregation of results by subgroups, required states to reach 100% student proficiency in math and English in grades 3-8 by 2014, and to reach annual Adequate Yearly Progress (AYP) targets between 2006 and 2014. AYP was composed of indicators for proficient performance and satisfactory participation on standardized assessments as well as reaching a target for an additional indicator (e.g., Attendance Rates, Graduation Rates) at both the primary and secondary levels. These requirements and federal pressure essentially compelled all states to adopt consequential accountability policies (Dee & Jacob, 2010; Hanushek & Raymond, 2005). NCLB provided states with the ability to seek approval at the federal level of their accountability plans, which provided the basic rules

and framework for statewide implementation of NCLB. Within these accountability plans and amendments to the accountability plans there was considerable variation, with many pointing out the numerous ways that states varied in their implementation of NCLB, such as differences in accountable subgroup size requirements and the definition of proficiency (Cullen & Reback, 2006; Davidson, et al., 2013; Dee & Jacob, 2011; Reback, 2008; Wei, 2012).

Variation in states' implementation of accountability stringency. Although NCLB required that all states achieve the 100% proficiency goal by 2014, the legislation allowed states considerable latitude in implementing its requirements, from the choice of statewide student assessments to the trajectory of proficiency standards to the use of exemption rules. One way that states have been able to alter the performance on standardized assessments is by selecting assessments that varied in their psychometric properties, content focus, and difficulty (Linn, 1999). Psychometrically-sound standardized assessments are a central element to any accountability policy and are key in gauging student performance. However, within this framework of psychometrically-sound assessments, states could choose assessments that varied widely in terms of length, difficulty, and composition of the tested material (De Mello, 2015; Dee & Jacob, 2010; Hanushek & Raymond, 2005). The variable quality and properties in assessments across states was an obvious concern, leading to uncertainty regarding the efficacy of the various state assessments in gauging states' ability to educate students. It is also made it difficult to compare student performance across states, as students across states were being tested with disparate assessments varying widely in difficulty and content (Linn, 1999). Psychometrically sound standardized assessments are a central element to any accountability policy and are key in gauging student performance. However, within this framework of psychometrically sound assessments, states could choose assessments that varied widely in terms

of length, difficulty, and composition of the tested material (De Mello, 2015; Dee & Jacob, 2010; Hanushek & Raymond, 2005). The variable quality and properties in assessments across states was an obvious concern, leading to an uncertainty as to the efficacy of the various state assessments in gauging the states' ability to educate students. It is also created a difficulty in comparing student performance across states, as students across states are being measured with disparate assessments varying widely in difficulty and content (Linn, 1999). There was extensive cross-state variation in the difficulty and content of the state-wide assessments, with many states choosing to employ "easier" tests, essentially inflating the proficiency scores obtained by using these tests. One group of studies (De Mello, 2011, 2015) investigated the cross-state variation in the difficulty of state-wide assessments utilizing the nationally representative National Assessment of Educational Progress (NAEP). Comparing the NAEP scores for each state to the state's standard for proficient performance generally showed an inordinate number of students passing their standardized exams but not the NAEP exams (De Mello, 2011, 2015).

Another source of variation in states' accountability policies was their choice of a threshold for the percentage of students who were required to be "proficient" in a given subject area (Krieg, 2008; Neal & Schanzenbach, 2010; Wong et al., 2017). For schools to meet federal AYP requirements, a minimum percentage of the students in each subgroup in a school must be deemed "proficient" in reading and math. However, states had discretion over the minimum proficiency requirements as long as they achieved the 100% proficiency goal by 2014 (Carey, 2007; Krieg, 2008). Under NCLB, states chose different trajectories for raising their proficiency requirements. For example, some states raised their proficiency thresholds incrementally over time; others held their thresholds constant for the first two years, and gradually increased them until 2014; still other states maintained their thresholds at near constant level for the first five to

six years, ratcheting their requirements dramatically in the latter period of NCLB (Krieg, 2008; Neal & Schanzenbach, 2010; Wong et al., 2017). An example of this behavior can be seen in Figure 1.

States could also adopt exemption rules that effectively lowered proficiency requirements for school subgroups. Prior research shows that states had latitude in employing exemption rules, which effectively lowered proficiency requirements for schools. These exemption rules were created to provide valid and reliable AYP calculations (Rogasa, 2003), but studies have shown that many states used these rules to reduce accountability standards (Davidson et al., 2013; Dee & Jacob, 2011; Wei, 2012). The exemption rules include the use of multiple years to average proficiency scores, the use of safe harbor calculations, and the use of confidence intervals (Davidson et al., 2013; Dee & Jacob, 2011; Wei, 2012; Wong et al., 2017). Multi-year averaging rules allowed states to let schools use student performance scores from one and/or two years prior to calculate an average score over those years if a subgroup, grade, or school fails to reach the required threshold in the current year; this essentially indicates that the current year's student performance is not representative of how well the school is performing over time. Safe Harbor rules required a 10% reduction in the percentage of non-proficient students within a subgroup from the prior year, effectively allowing subgroups to be proficient even when they fell well-below the minimum proficiency threshold. Finally, confidence intervals accounted for the possibility of random error in the process of calculating student performance, as well as allowed states to pass subgroups that were near the proficiency threshold. Given the number of students in a group, the rule constructs a plus/minus band around the proficiency threshold effectively extending the cutoff to several percentage points below the initial state threshold.

A small number of states also utilized performance index models in their determination of AYP, a strategy that essentially boosts aggregate school performance numbers (Wong et al., 2017). The use of a performance index model places weights across various performance levels and does not only count the percentage above proficient; instead, it provides partial credit to students who did not reach the proficiency threshold. For example, the state of Alabama placed a weight of 1.0 on students who were at or above proficiency, a weight of .5 for those partially meeting standards, and a weight of 0.0 for those not meeting standards. While some states provided half credit to those who scored around a basic level of performance, other states such as Massachusetts and South Carolina utilized a more detailed tier structure that further divided up the possible performance levels.

No Child Left Behind waivers and ESSA. In the early-2010s, states were given the option of applying for waivers from some of the more stringent parts of NCLB, such as the need to attain 100% proficiency in math and reading. This created an even greater amount of variation in how states enacted NCLB and was shown to help improve math and reading achievement (Bonilla & Dee, 2017). In late-2015, the Every Student Succeeds Act (ESSA) was signed into law as a replacement for No Child Left Behind. ESSA moves away from the extensive involvement of federal government and moves to a more flexible, state and locality driven policy (Klein, 2016). The ESSA provides guidelines and a framework for holding schools accountable, but there is a slightly more progress-focused approach. At the same time, under ESSA states are still required to administer reading and math tests to all students in grades 3-8.

Teachers' Roles and Responses in Organizational Systems

Teachers have a central role in the success of accountability reform efforts. In this section, I examine general factors that are related to teacher employment outcomes, daily work

experiences, and interest in the teaching profession, and consider how these findings inform our understanding of teachers' responses to accountability policies.

Within any professional field, there is a natural level of employee turnover year after year; whether through voluntary or involuntary means, there is an expected amount of turnover in any place of employment. There are also aging employees entering retirement, employees moving into more administrative and supervisory positions, and individuals leaving the field to join another field or who are removed from their position. However, within the field of teaching, there is evidence that teachers move out of the system at a much higher rate than most fields, and even on par with jobs associated with high danger situations, such as police officers (Ingersoll, Merrill, & May, 2016). The Schools and Staffing Survey (SASS) was conducted between 1990 and 2012 and renamed the National Teacher and Principal Survey NTPS in 2016. Analyses using SASS show that teachers left their positions at a range of rates ranging from 13.1% to 16.7% a year, with less than 15% of this attrition due to retirement. A vast majority of attrition was due to pre-retirement causes and the rates were higher for first- and second-year teachers (Kelly & Northrop, 2015). While employment and teacher turnover has generally been stable over time, the trend has become more volatile in recent years (Ingersoll, Merrill, & Stuckey, 2014). Before the 2017-2018 school year, there was an estimated teacher shortage of 110,000 teachers (Sutcher, Darling-Hammond, & Carver-Thomas, 2016), which is a sharp increase from no shortage before the 2012-2013 school year. Similarly, the number of schools reporting any vacancy increased from 67.2% to 78.8% from 2012 to 2016. Research has shown that this teacher shortage is largely a result of high teacher attrition and a declining pool of applicants ((Darling-Hammond et al., 2017; Sutcher et al., 2016).

It is worth noting that at times, a teacher leaving a position is not an overtly negative thing, with low-performing or less motivated teachers being the ones who tend leave the profession (Adnott, Dee, Katz, & Wyckoff, 2017). However, low-quality teachers leaving the field creates a void that needs to be filled by another teacher, which can lead to costly teacher searches, recruitment strategies, development costs, and changes in the composition of teachers within schools (Ingersoll & Perda, 2010). This void can also lead to staffing shortages within the schools that have difficulty recruiting new teachers, leading to issues such as poor student performance and a decrease in teacher morale. In monetary terms, teacher turnover costs school districts up to an estimated 2.2 billion dollars a year (Haynes, 2014).

Teacher Responses to Accountability Policies

Prior research on NCLB and teacher outcomes has suggested that teachers respond to accountability pressure in a number of ways that are both positive and negative. In particular, this literature has examined how teachers have responded to the pressures of high-stakes testing and teacher evaluation systems. In this section, I examine teachers' responses to accountability pressures, including how these reform efforts have changed their perceptions of their profession and job security, as well as their attitudes towards teaching. I also examine less desirable responses to accountability pressures, including tailoring instruction towards specific students and subjects, teaching to the test, and manipulating students' responses on assessment sheets. Finally, I examine teachers' reports of their perceived control in the classroom, how the school environment has changed under NCLB, and how teachers are moving in, out, and within schools due to accountability requirements.

Perceptions and attitudes of teachers. Prior research has examined how accountability policies such as NCLB have impacted teacher job perception and security, as well as teacher

attitudes towards their profession and their overall well-being. These studies have shown mixed results, although most studies that have examined NCLB's impact have suggested negative responses from teachers in terms of their attitudes, perceptions, and feelings of job security (see, for example, Clotfelter et al., 2004; Grissom et al., 2014; Rubin, 2011). While these studies provide context for teachers facing consequential accountability, they do not show how teachers respond to differential levels of accountability stringency under NCLB. This study looks at many of the same outcomes as the following research studies, but it investigates how teachers respond to higher or lower levels of accountability stringency.

One key outcome is how teachers feel about the teaching profession and how they perceive others to feel about the teaching profession (Clotfelter et al., 2004; Feng & Sass, 2008; Rubin, 2011). Many teachers have realized a shift from valuing their ability to engage a classroom and teaching the "whole" student, to valuing how well they prepare students to perform successfully on standardized assessments. Rubin (2011) found that during NCLB, English Language Arts Teachers felt "victimized" because of the additional pressure placed upon them by the various accountability policy rules and requirements. The shift from perceiving a quality teacher as one who uses creative problem-solving and engaging activities to reach students to a teacher who leads students to performing well on standardized assessments led many teachers to be disheartened with their chosen profession (Rubin, 2011). Teachers that felt this shift in perception and the pressure of using standardized assessments as an evaluative measure had greater levels of stress and anxiety than those who did not perceive the shift or face evaluative standardized assessments (Haladyna, Haas, & Allison, 1998; Mulvenson, Stegman, & Ritter, 2005).

Some studies have also found that consequential accountability policies, specifically NCLB, have affected how secure teachers felt within their profession. For example, Reback, Rockoff, and Schwartz (2014) observed that teachers who were in schools near the Annual Measurable Objective felt increased pressure, possibly due to the threat of failing to reach Adequate Yearly Progress. In survey responses, teachers report that they felt overburdened by the pressures of NCLB and that they had a critical view of the policy (Center for Education Policy [CEP], 2005; Sunderman, Tracey, Kim, & Orfield, 2004). These surveys also indicated that staff morale declined under NCLB because of the emphasis on standardized testing (Byrd-Blake et al., 2010; CEP, 2005). Staff morale also decreased when schools failed to meet AYP requirements, or were placed under corrective measures (Santoro, 2011).

Although most studies mentioned above indicate a negative association between job perceptions, attitudes, and job security, not all studies investigating the tie between teachers and accountability policy were wholly negative. Grissom, Nicholson-Crotty, and Harrington (2014) found that although there were some negative impacts of NCLB on teachers, there were also some positive or mixed outcomes. By looking at states that previously had consequential accountability systems prior to NCLB and following the trends in teacher satisfaction and attitudes, and other outcomes as gauged through the Schools and Staffing Survey from 1994 to 2008, they reported a positive impact on the amount of autonomy and control teachers felt within their classroom and on the amount of support they received from the administration. However, this study also illustrated a negative impact on teacher cooperation and little to no impact on teacher job satisfaction or commitment to the job. Other studies (Hamilton et al., 2007; Murnane & Papay, 2010) echoed the positive findings of the Grissom, Nicholson-Crotty, and Harrington

study and found that teachers perceived that they had more control within the classroom and more direct guidance in regards to expectations of students and student learning.

Changes in teaching practices. Given the central role that teachers play in improving student performance, prior research has examined whether and how teachers changed their instructional practices to meet accountability requirements. These studies found that many teachers responded to accountability pressures by increasing efforts to reach specific students, and by focusing on specific tested subjects (Booher-Jennings, 2005; Figlio & Rouse, 2006; Jacob, 2005; Neal & Schanzenbach, 2010; Reback, 2008). Neal and Schanzenbach (2010) utilized the varying levels of proficiency thresholds in Chicago during the pre-NCLB era and discovered that teachers devoted differential time to students of different ability levels based upon these proficiency thresholds. They reported that teachers were consistently placing more focus on students who were expected to score near the proficiency threshold. Teachers also shifted their focus to providing increase instructional time for tested subjects, and also focused their materials to address content that was expected to be covered by the test (Booher-Jennings, 2005; Dee, Jacob, & Schwartz, 2013; Figlio & Rouse, 2006; Jacob, 2005; Neal & Schanzenbach, 2010; Reback, 2008). Many teachers took it upon themselves to ensure that they focused on preparing students for the standardized assessments, which affected how they were evaluated as a teacher. In addition, teachers within schools near proficiency thresholds worked longer hours (Reback, Rockoff, & Schwartz, 2014).

Finally, teachers also engaged in illegal practices for ensuring students met accountability standards. For example, teachers in Atlanta Public Schools (Blinder, 2015) and other districts (Amrein-Beardsley, Berliner, & Rideau, 2010) engaged in large- and small-scale cheating strategies to increase students' performance on standardized assessments. The teachers in Atlanta

were correcting student answers and utilizing other techniques, thereby creating statistically abnormal gains in assessment scores from one year to the next. Past research (Jacob, 2005; Jacob & Levitt, 2003) indicates that the Atlanta example is not the only case of teachers cheating to improve student test scores. Jacob and Leavitt (2003) conducted a study looking at the impact of the introduction of high-stakes assessments into the accountability regime in Chicago during the pre-NCLB accountability era. Utilizing an algorithm they developed to detect cheating, they found that cheating occurred in 4 to 5 percent of elementary school classrooms after the introduction of the new high-stakes, consequential accountability policy. The covertness and general riskiness of cheating make it quite difficult to detect; thus. the research on cheating is not as abundant as that on other strategies and responses utilized by teachers and schools.

Teacher staffing and turnover. During the normal ebb and flow of the teacher labor markets, teachers tend to move in and out of the teaching profession with relative ease compared to other professions. Many teachers stop teaching for a period of time for various reasons and then move back into the profession. The focus of this study, however, is on how accountability pressure may hasten or delay this process. Prior research suggests that accountability pressures have altered teachers' decisions to stay and leave schools, as well as to leave the teaching profession altogether (Dee & Jacob, 2010; Feng, Figlio, & Sass, 2010). Feng, Figlio, and Sass (2010) observed the responses by teachers to schools' accountability performance by utilizing an exogenous sudden change in policy that occurred in Florida in 2002. In Florida, from 1999 to 2001, schools were graded from A to F based on how students performed on reading, math, and writing assessments. However, in 2002 the grading system changed to be more in line with current accountability standards and caused over half the schools to receive a different grade than would have been expected using the earlier system.

As this change was unexpected, it allowed the researchers to causally investigate if a teacher was more or less likely to leave schools that were suddenly shifted to a lower or higher accountability grade. Feng, Figlio, and Sass (2010) found that schools that were shifted down to a lower accountability grade saw an increase in the likelihood that teachers would leave those schools, where the opposite occurred in schools that were shifted upward. This finding, and the findings of other similar studies on consequential accountability policies (Clotfelter et al., 2004; Feng & Sass, 2008) indicated that teachers were attempting to set themselves up in the best academic situation by choosing to go to better-performing schools. However, these results were not replicated in a comparable study that used a differences-in-differences approach. These researchers found no impact of NCLB on teachers voluntarily leaving schools (Sun, Saultz, & Ye, 2016). Sun, Saultz, and Ye (2016) reported that teachers were involuntarily transferred at greater rates during the early years of NCLB, and Ingersoll, Merrill, and May (2016) found that poor-performing schools were much more likely to have issues with teacher retention and staffing problems. Although involuntary transfers were undesirable for teachers, they may have benefited the school if it provided opportunities to remediate low-performing teachers (Hanushek & Rivkin, 2010).

One other response, although not as thoroughly researched or studied, is the sorting of teachers within schools, essentially shifting teachers either from tested to untested subjects or grades or shifting teachers from untested to tested subjects or grades. This sorting within schools occurs if school administrators move a 2nd-grade teacher (which is an untested grade) up to 3rd grade (i.e., a tested grade) to improve standardized assessment performance and student achievement. Sorting behavior has been seen in schools as a response to accountability policy, with less credentialed teachers shifting down to the grades that are not held accountable and higher credentialed teachers shifting up to the grades that are held accountable (Fuller & Ladd, 2013). Sorting within schools and shifting higher credentialed teachers to tested grades and subjects may appear to be an advantageous strategy. However, it is unknown whether the shifting of higher-quality teachers to tested grades will be mitigated by the loss of these teachers in younger grades.

Potential Moderators of State Accountability Stringency on Teacher Outcomes

One of the key moderators of state accountability stringency on teacher outcomes is the presence or lack of presence of consequential accountability policies in states prior to No Child Left Behind. Hanushek and Raymond (2005) highlight that during NCLB all states have consequential accountability, whereas, this was not the case before NCLB. Research done by Hanushek and Raymond, among others (Dee & Jacob, 2011; Carnoy & Loeb, 2002), illustrates that states with previous consequential accountability policy had less of a lift in reaching the goals set forth by NCLB since they already had a similar policy in place. This study will take this into account, but it will also take it a step further by estimating the strength of previous accountability policy, therefore, demonstrating that not all states with previous accountability policy had the same level of sanctions, provisions, and requirements. This extension not only

helps to account for the amount of shock a state faces when NCLB is enacted, but it also allows the study to investigate if states with previous accountability policies start to get exasperated with the stress put on them by consequential accountability policies.

Prior organizational theories (e.g., Ingersoll, 2001; Karasek & Thorell, 1990) suggest that there may be potential moderators for how teachers respond differentially to accountability stringency. For example, contemporary education theory (Ingersoll, 2001) suggests that demographic patterns, such as age and tenure of the teacher, may influence teacher responses under accountability policies. Large policy changes may nudge teachers who are on the verge of retiring into retirement, leaving many openings for new and less experienced teachers. With the idea that the "Baby Boomer" generation was moving into retirement age, this could be a factor in this study. The departure of experienced, high-quality teachers from schools that already are at a higher risk of failing to reach performance thresholds can leave such schools with a large percentage of young and inexperienced teachers who may find it difficult to improve student performance. Research indicates that young, less-credentialed teachers often are assigned more disadvantaged students and students who have issues with discipline, compounding the difficulty in improving student learning and achievement (Boyd et al., 2005; Clotfelter, Ladd, & Vigdor, 2005, 2006, 2011; Feng, 2010; Krei, 1998; Lankford, Loeb, & Wyckoff 2002; Rice, 2013). As a result, this dissertation examines the heterogeneity of teacher responses to accountability stringency by differences in state composition of teacher characteristics such as age and tenure.

Along the same lines as demographic characteristics of teachers, an organizational perspective (Halaby & Weakliem, 1989; Hom & Griffeth, 1995; Mueller & Price, 1990; Price 1977, 1989) suggests that existing school conditions may also mitigate teacher responses to accountability pressures. Teachers who work in schools that are at-risk for failure, or have highly

disadvantaged populations, may feel greater pressures and be likely to leave their positions than teachers in higher-performing and less-disadvantaged schools. Another factor related to existing school conditions are the economic and demographic makeup of students within the school, with those schools with a high percentage of low-income and non-white students losing experienced and higher-quality teachers at a faster rate (Hanushek, Kain, & Rivkin 2004; Scafidi, Sjoquist, & Stinebrickner 2007). Schools with a high percentage of low-income students also attract a higher percentage of minority teachers' although minority have been entering the teaching profession at a high rate, they have a higher rate of teacher turnover than white teachers and feel a general lack of control and autonomy within the classroom (Ingersoll & May, 2011). This dissertation examines the responses of teachers to accountability stringency based on school conditions.

Finally, the Demand-Control-Support (DCS) theory suggests that the working conditions within the school, accounting for such thing sas class sizes, leadership, and salaries, can be highly associated with teacher mobility and attrition (Karasek & Thorell, 1990),. The DCS model argues that a position with high demands, but low support and an inability to influence organizational decisions leads to higher amounts of stress and other related issues. Furthermore, teachers' reported control within the classroom, perceptions of working conditions and perceptions of support from colleagues and leadership have been linked with teacher turnover and employment outcomes (Grissom et al., 2014; Ladd, 2011).

Contribution of Present Study

This dissertation contributes to the research on the impact of accountability policies on teacher outcomes by utilizing a measure of implementation stringency over the many years of No Child Left Behind. Several of the studies previously discussed with robust evidence investigated the introduction of new accountability policies, not changes in accountability stringency. Many

of the studies used a similar approach investigating how the presence or lack of presence of an accountability policy before NCLB impacted various stakeholders, essentially treating NCLB as a shock to the system. Dee and Jacob (2010, 2011) utilized this approach when gauging the impact of NCLB on various student, school, and teacher outcomes and this study has been followed by other similar studies (e.g., Grissom, Nicholson-Crotty, & Harrington, 2014). The basic approach follows the differences-in-differences framework that investigated outcomes across treatment and control groups pre- and post-NCLB. The treatment group in this framework included the states that did not have a consequential accountability policy in place prior to the start of NCLB, while the control group included the states that did have consequential accountability policy in place prior to the start of NCLB. This leads to the notion that NCLB is a shock to a state that did not have a previous accountability system, thereby creating the assumption that No Child Left Behind would cause greater impact on these treated states.

This strategy for exploring the impact of No Child Left Behind is sound and innovative, however it does not address changes in stringency over time. My study is important because it addresses these changes in stringency over time, instead of as only a sudden shock to the system and it also helps to inform policymakers in regards to whether or not the benefits of a new accountability system can be sustained long term. Instead of only investigating how the introduction of a new accountability policy impacts teachers, this study investigates how teachers, who were on the frontlines of meeting increasing proficiency requirements under NCLB, reacted to the changes. Using the previously discussed stringency measure this study delves into how changes in stringency impact the overall experience of being a teacher during NCLB. This study will also utilize a measure for pre-NCLB strength in order to account for differential shocks to the system, where states without previous accountability policies or weak

previous accountability policies may feel more of an impact of a new policy than those with

strong previous consequential accountability policies.

METHODOLOGY

This study uses simulated AYP school failure rates and two-way fixed effect models to examine the causal linkages between states' accountability decisions and teacher outcomes. The study contributes to the existing accountability literature by utilizing a continuous measure for accountability stringency, applying this measure across many states and years, and investigating the causal impact of stringency on several teacher outcomes. In this chapter, I describe the research design for addressing this question and plausible threats to validity. I then describe the independent and dependent measures used in this study, including an overview of how the implementation stringency measure was created. Finally, I summarize the sensitivity tests and robustness check utilized for validity threats to the research design.

Research Design

To examine the impact of increased accountability stringency on teacher outcomes, in an ideal world, we would randomly assign states with differing levels of accountability stringency and observe changes in teacher workforce outcomes. However, random assignment of accountability stringency is infeasible under NCLB, so it is essential to utilize a quasi-experimental design that allows for a causal analysis of the effect accountability stringency. Fortunately, the implementation of NCLB provides a natural experiment by which we can investigate the impact of accountability stringency on teacher outcomes. To accomplish this task, I utilize a state-by-year fixed effects approach of teacher outcomes on an annual quantitative measure of the states' accountability stringency rates. The two-way fixed effects approach is a more general version of a differences-in-differences design, which typically utilizes a binary treatment variable and two separate data points before and after an intervention. Prior evaluations of NCLB utilized a differences-in-differences design with dichotomized treatment variable

indicating either the presence or absence of pre-NCLB consequential accountability. This study, however, uses a continuous treatment variable based upon variation in accountability stringency during NCLB. The generalized approach of this study allows for the use of a more flexible treatment indicator and a more adaptable framework than the simple differences-in-differences approach.

Threats to Validity

The study's approach provides the advantage of being able to utilize a flexible model and a flexible treatment indicator, however, there are still a few assumptions required for the research design to yield valid treatment effects. One key assumption is the common trend assumption. This requirement states that in the absence of treatment, the average change in the outcome would have been equal between all groups. In this instance, this would mean that teacher outcomes across the states would have had followed the same trend in the absence of NCLB. The two-way fixed effects approach aids in relaxing this assumption as it allows me to utilize a series of data points both before and after the implementation of NCLB, providing a more precise and reliable picture of policy impacts than the typical DID approach. Analytic samples that include several states and years provide a means to ease and check the common trends assumption by allowing for the use of state by year interaction terms to account for varying unobservable trends.

The common trend assumption could be violated in multiple ways. For example, regression to the mean is one possible limitation within this study, where teacher outcomes collected for the study are extreme outliers because of some sort of chance event. In the context of this study, regression to the mean could follow an Ashenfelter's Dip-type phenomenon, where a dip or trend occurs right before treatment in states that have extreme stringency values. States with higher levels of teacher turnover directly before the implementation of NCLB may react by

enacting an accountability policy that is highly stringent. The pre-treatment spike in teacher

turnover can then lead to an estimate showing that highly stringent policies lead to a decline in

teacher turnover, when the estimate may realistically only be picking up a natural regression to

the mean. To check for an Ashenfelter's dip, I utilize lead treatment indicators to measure the

change in slope for states that are about to implement NCLB and ensure that lead treatment

indicators have no impact.

Another limitation and possible source of bias in this study are the history effects caused

by events other than the implementation of NCLB that could explain the changes in teacher

outcomes. For instance, the recession of 2008 could have caused a large number of teachers to

stay within their position in order to maintain steady income or it could have caused a dip in the

tax revenue going to schools. If the recession affected all of the states the same, then year fixed

effects would address this; however, if the recession differentially impacted the states than

history effects would be an issue. This could result in estimates that point to NCLB impacting

teacher turnover instead of properly attributing impact to the recession. To attempt to address

this limitation, I utilize a comparison group composed of state employees that also would face

many of the same historical factors and events as teachers. Also, to gauge the expected level of

impact of the recession I utilize a variable that measures the severity of the recession within each

state.

Another threat to validity is that compositional differences of teachers and students

within states may have occurred between 2003 and 2011. This is a plausible validity threat if for

example, a state has a tremendous influx of new teachers entering the field while older teachers

are retiring. This would cause a sudden shift in the composition of teachers and the average level

of experience within the state. To address this threat, I control for time varying factors across the

states, such as teacher population characteristics, state unemployment rates, population demographic characteristics, and more. The teacher population characteristics include such things as the teacher's age, race, and others that will control for changes in the composition of teachers. Unemployment rates and population demographics characteristics are also taken into account in order to capture changes in income levels, average age, education level, and other characteristics that may impact stringency and teacher outcomes. It's worth noting that one benefit of the stringency measure is that it is independent of population characteristics of the state. That is, it calculates the simulated failure rate for a fixed sample of students and schools and accounts for the differences in accountability policy implementation, but not for the performance level or characteristics of the state's students.

The history threats and composition change threats can be classified under the threat of omitted variable bias. Omitted variables are variables that confound and separately affect both the independent and dependent variables and also are correlated with the error term. These variables, when left out of the model, would create bias and cause the over or under estimation of the treatment effect. Omitted variables in this study could include a factor that indicates a state's ability to motivate teachers to stay within their position, a state's attitude toward accountability policy, or the ability of a state to retain teachers of a certain mindset. This factor could affect how states choose their accountability policy implementation rules and impact various teacher outcomes. This threat is addressed through the use of the general, two-way effects strategy that addresses both time invariant factors within states (state effects) and factors that affect all states at the same time (year effects).

Along the same lines as history threats is the possibility that the Great Recession of 2008 may have had a major impact on both how stringent states were in their accountability policies

and also in how teachers reacted to changes in accountability policy. The Great Recession had a major impact on the economy of the United States and had major ramifications in many fields (Kalleberg & Von Wachter, 2018), including the education sector (Evans, Schwab, & Wagner, 2019). Results from the various surveys this study investigates may be impacted by the way teachers and states were dealing with the recession. To address the threat of the impact of the Recession on the analysis, this study utilizes a measure that quantifies this impact for each state and includes it in the analytic models.

One final threat to acknowledge is the attrition of teachers between surveys. Since this study looks across several years and across several survey cycles, it is highly likely that the population of the teachers taking the surveys will be different across the years. Changes in survey results could become misleading and wrongly attributed to variation in accountability policy. Addressing the issue of attrition helps to properly ascribe changes in results to either variation in stringency or to underlying changes in teacher demographics. The threat of attrition is addressed in the study by analyzing the method by which the surveys were sampled and by statistically investigating whether the underlying teacher characteristics significantly varied across the surveys.

Sample and Context

The sample for this study consists of 49 states and the District of Columbia from 2003 to 2012, as full implementation of NCLB occurred during the 2002-2003 school year. The lone state being left out of the sample completely is Ohio which utilized a growth model that I was not able to incorporate into the stringency measure for this study. While the application of waivers for NCLB began in the 2011-2012 school year, 2012-2013 marked the beginning of the full implementation of waivers. This period is not included in the analysis sample as the focus of

implementation shifts from the state level down to the school and district level. For example, waivers allowed for differential AMO thresholds within the state, making it difficult to create a single stringency measure for the entire state that is representative of the experience for the teachers. Further discussion, exploration, and description on the sample of states and the samples from the key outcomes will be a part of the proceeding sections that cover the creation of the stringency measure and the multitude of teacher experience outcomes.

Independent Variable: Stringency

To capture and study the impact of differential accountability policies on teachers during NCLB I use a stringency measure created by Wong et al. (2017), which encapsulates nearly all of the variation across the states and years during NCLB. This measure encompasses the state rules and the differential policies across states and years and provides an estimate for how difficult it is to reach AYP given a state's accountability implementation within a given year. The result is a state-by-year level dataset summarizing AYP failure rates for a fixed sample of schools.

The stringency measure is available for all 50 states and for the District of Columbia and goes through a majority of the years that NCLB was implemented (2003-2011). The key factor to note about the measure is that it is not dependent on state population characteristics as a fixed sample of schools was used that did not change across states and time periods. Therefore, the variation in stringency is a result of the differential implementation of accountability rules during NCLB. The stringency measure encompasses all of the decision rules in regards to the implementation of NCLB and their incorporation is detailed in the appendix.

Fixed sample characteristics. Two challenges arise in the effort to assess states' implementation stringency, including how to create a measure independent of each state's

population characteristics and how to account for the differences in test difficulties across the state. To address the first concern, the stringency measure uses a fixed sample of schools made up of a national dataset of students who took the National Assessment of Educational Progress (NAEP) assessments. This sample only consists of 4th and 8th grade students, but it is highly representative of the national landscape and its characteristics and reflects sizeable variation in subgroup sizes and performance to compare against the variable state rules. Table 1 illustrates the overall demographic characteristics of the NAEP sample.

This fixed sample of schools allows for changes in the stringency measure to occur because of changes in the accountability rules by state and year and not by the characteristics of an ever-changing population of students. If the measure utilized samples from within each state, it would be possible that any changes in stringency would be because of some shifts in the performance and/or characteristics of the student body.

Addressing differences in test difficulty. To address the concern regarding differential statewide assessment difficulty, the measure utilizes a method of mapping that links state assessments to NAEP assessments. Taylor et al (2010) noted the large amount of variation in test difficulty across the states and compared the performance of students on their statewide assessments and their performance on the nationwide NAEP assessment. Using a similar method, the NCES NAEP State Mapping Project created a series of NAEP equivalent score cutoffs that map scores from the state assessments onto NAEP scale scores for 4th and 8th grade students in 2003, 2005, 2007, 2009, and 2011. To create the NAEP equivalent cutoffs, a mapping links the percentage of students in a state achieving proficiency on their standardized assessment to the NAEP score that has the same percentage of students scoring above that value. For example, if 34% of the students in a state achieve proficiency on a math assessment that matches to the

NAEP score that has 34% of the students scoring above that value, this creates the equivalent

NAEP score. This equivalent NAEP cutoff score helps to show test difficulty with higher NAEP

equivalent scores illustrating a more difficult statewide assessment and lower NAEP equivalent

scores representing an easier statewide assessment and it also illustrates changes in the chosen

statewide assessment. NAEP equivalent cutoffs were not available for every state and year so the

stringency measure uses interpolated NAEP equivalent cutoff scores for years in which

information was not available. The full list of NAEP equivalent cutoffs is available in the

appendix.

Although the stringency measure attempts to fully incorporate each of the main features

of the state's rules, not all actual accountability rules could be fully implemented. Table 2

categorizes the states into different groups based upon how closely the AYP calculator is able to

approximate that state's rules. The first column illustrates those states (25 states and DC) that are

high-fidelity states and that are a close representation of the actual rules for those states. The

second column shows those states (24 states) that have rules that needed to be approximated and

implemented in the AYP calculator with partial fidelity. The partial fidelity states had

accountability rules that involved modified proficiency and exemption rules, the use of

performance/proficiency indices, and/or additional academic indicator requirements that used

science or writing assessments. The modifications to these rules did not appear to greatly impact

the measure as was shown in our validation efforts. The one state (Ohio) that has low

implementation fidelity was unable to be appropriately approximated because of the use of

complex growth model rules that I was unable to incorporate into the measure.

In order to assess whether the calculator properly implemented state accountability rules,

we utilized several empirical validation exercises across several states. To gauge the validity of

the stringency measure, we obtained data on the population of schools in Texas, Pennsylvania, Maryland, California, New York, North Carolina, and Ohio for two separate time periods each. We utilized the actual test scores for the given year and state and ran them through the calculator using the implementation rules for that given year and state. We then compared the states' actual failure rates to the calculated failure rates from the AYP calculator. If the predicted failure rates were close to identical to those of the actual failure rates, then the calculator should be correctly accounting for the rules. Table 3 shows the comparisons between the predicted and actual failure rates for the given states and years. In general, this table illustrates a close relationship between the predicted and the actual failure rates. In some cases, input was lacking for specific rules or populations within the states, which could be a cause for discrepancy. The one state that had a clear discrepancy was Ohio, which is the reason why it was not included in the analysis sample.

Dependent Variables: Teacher Experience Outcomes

This study utilizes a number of datasets and sources to investigate the various teacher experience outcomes, including the Schools and Staffing Survey, the Teacher Follow-Up Survey, the Freshman Survey, the College Senior Survey, the Bureau of Labor Statistics, the Occupational Employment Statistics, the Current Population Survey, and the Common Core of Data (CCD). The following sections provide an outline of these data sources and the specific outcomes that come from these data sources.

Schools and Staffing Survey/Teacher Follow-Up Survey. The School and Staffing Survey (SASS) and the Teacher Follow-up Survey (TFS) provides outcome information about how teachers are moving in and out of the profession, their perception of control in the classroom, their satisfaction with their career, and their perception of support from the administration and other colleagues. These surveys consider of nationally representative samples

collected and designed by the National Center for Educational Statistics of the U.S. Department of Education with assistance from U.S. Census Bureau, for a majority of the teacher labor and teacher perception outcomes. Since 1988, the SASS has been conducted every several years and surveys teachers and administrators about specific issues and mechanisms related to their schools, teachers, families, students and administrators. NCES uses a stratified probability sample design selecting teachers and schools across the nation into the SASS and the SASS sample comprises over 40,000 to 50,000 teachers that teach in either public or private schools. Schools are stratified and sampled and then teachers within those schools and stratified and sampled based on a number of characteristics. The teacher-focused SASS asks teachers several categories of questions ranging from their thoughts on professional development opportunity, control within the classrooms, problems with students in their school, and many other questions that help to round out the teacher experience. The SASS has administered surveys in 1988, 1991, 1994, 2000, 2004, 2008, and 2012 and has also regularly conducted the Teacher Follow-up Survey the following year after the SASS.

To obtain the sample for the TFS, school administrators provide responses to the Teacher Status Form to indicate the employment status of each of the teachers surveyed in the previous year's School and Staffing Survey. NCES then creates a stratified sample of the teachers from the SASS based upon groups of teachers that stay with the same school (stayers), who left the profession altogether (leavers), or who move to a different school (movers). NCES then provides the Teacher Follow-up Survey, Current Teacher edition to teachers that stay within the same school or move to a different school and the Teacher Follow-up Survey, Former Teacher edition to those teachers that are no longer teaching. The Current Teacher Survey asks some similar questions to those found in the SASS, but also goes deeper into the reasons why a teacher chose

to stay in the same school or move to another school and the Former Teacher Survey provides

depth into the reasons why teachers leave the profession and what drove them to move in another

direction.

The main sample for this study contains only full-time, public school teachers, as private

school teachers do not face the standards and rules of NCLB and part-time teachers would feel

the impact in a vastly different and unique way. The sample sizes for the SASS are seen in the

Table 4, with a state-by-state table looking at the number of teachers in each state taking the

SASS is included in the appendix. Private school teachers from this dataset will be used as a

comparison group in order to compare groups that were facing similar statewide policies without

facing the pressure of NCLB.

The following sections describe the outcomes retrieved from the School and Staffing

Survey and the Teacher Follow-up Survey.

Teacher employment outcomes. The teacher employment outcomes for this study

encompass the number and percentage of stayers, leavers, and movers within each of the states.

The leaver and mover outcomes are broken further down into smaller subcategories. The leaver

category is broken down into whether a teacher left teaching voluntarily and why they left

(retirement, better career choices, among others) and whether a teacher left involuntarily

(contract terminated). The mover category is similarly broken down by looking at teachers that

moved schools voluntarily or those teachers that moved schools involuntarily.

Teacher attitudes and perceptions. Each measure of teacher attitudes and perceptions

provides a composite of several different survey questions, each of which are linked to in

Appendix D. The variable "control in the classroom" measures (alpha = .76) how much control a

teacher says that they have to make decisions on their own within their classroom, and teachers'

attitudes towards their autonomy for making decisions. The measure of job and salary satisfaction (alpha = .62) indicates whether teachers are satisfied with being a teacher at their school, whether they think other teachers are satisfied at the school, if things are going well at the school, and if they feel they are receiving fair compensation. Another aspect of the teacher experience that is measured is the teacher perception of the amount of support received from the school's administration (alpha = .83). This measure includes the amount of encouragement received from the administration, the clarity in the mission and rules within the school, and recognition from the administrator. The long-term outlook or career satisfaction measure (alpha = .57) looks at how teachers feel about their future within the teaching profession and how their choice to become a teacher. This measure combines two questions regarding whether a teacher would choose a teaching career again if they had to do all everything over, and how long they plan on staying within the teaching profession. The measure has a relatively low amount of reliability, however, it is an interesting and informative indicator for how teachers are feeling about their future in teaching. However, with the low level of reliability, the consistency of the outcome measure is an issue and the results gained within this study may only be picking up on the inconsistency of the measure. Another outcome that I analyzed in the study is the support received from colleagues (alpha = .74), specifically in regards to how teachers enforce rules and how the cooperate with each other. Additionally the burnout and career satisfaction (alpha = .80) measures are combined as well as the burnout, career, and job satisfaction measures (alpha = .83).

The Freshman Survey/College Senior Survey. Another main source of data for the outcomes in this study are the Freshman Survey and College Senior Survey developed by Higher Education Research Institute (HERI) and the Cooperative Institutional Research Program (CIRP)

from the University of California Los Angeles. The Freshman Survey is an entrance survey for first-year college students and provides information on possible career choices, planned field of study, expectations of college, preparedness for college, as well as basic student background characteristics. The College Senior Survey is either a standalone exit exam for outgoing seniors, or a longitudinal follow-up measure of students who took part in the Freshman Survey. The College Survey provides researchers and institutions with information on the student's experience during college, interactions with peers and faculty, future degree and career plans, and their overall satisfaction within the college.

Interest in the teaching profession. The outcomes analyzed from the Freshman Survey and College Senior Survey align with the idea of attracting new teachers to teaching and into the education system. These outcomes come in the form of the number of students looking to become education majors, to have a career in teaching, and to go on to an education master's degree after receiving an undergraduate degree, all of which illustrate how attractive teaching appears to new college students and how likely these students are to actually pursue a career in teaching.

Pre-NCLB accountability strength. To create a measure for the strength of pre-NCLB accountability, I built off of the work from Dee and Jacob (2011), among others (Carnoy and Loeb, 2002; Hanushek & Raymond, 2005; Lee and Wong, 2004). Dee and Jacob (2011) utilized these sources, as well as, many other sources such as data from state Department of Education web sites and conversations with state officials, to determine the states that did or did not have pre-NCLB consequential accountability. They confirmed the results of previous work done by Hanushek and Raymond (2005) to identify the states that did have pre-NCLB consequential accountability policy. To take this a step further, this study utilizes the work done through

Carnoy and Loeb (2002) and Lee and Wong (2004) to provide a measure for the strength of the pre-NCLB consequential accountability policy. Carnoy and Loeb (2002) created an accountability index from 1 to 5 where they researched whether a state required student testing, performance reporting, had sanctions or rewards and whether students needed to pass an exit to graduate. For this study, the more of these requirements that a state had in place, the higher the pre-NCLB accountability strength. In addition to the Carnoy and Loeb index, I utilized a measure from Lee and Wong (2004) that investigated similar requirements to place pre-NCLB strength on a scale from weak to moderate. States with weak strength did not provide direct incentives to schools, while states with strong accountability had all of these policies in place. From here, I created a measure that took into account the overall picture created by these measures. When there were no discrepancies across the measures, I utilized that score. If there were discrepancies, I took the majority decision and chose that as the strength for that state. For example, if a state was deemed by two studies to have strong previous accountability policy, but one study deemed it to have moderate, I placed the state into the strong previous accountability policy category. The detailed coding of these states is shown in Table 18 in Appendix F. By interacting a measure for previous accountability status with stringency variable, I am able to measure the differential impact of stringency on states that did or did not have previous accountability policy. States that did not have previous accountability policy may feel a larger increase in pressure from the institution of a new consequential accountability policy than those that did have a previous policy. By creating a measure for strength, I can also explore if states with a lower level of consequential accountability policy feel this shock differently than those states with more difficult previous accountability policies.

Other data sources and control variables. This study utilizes several federal level data sources to capture various control variables, which include student to teacher ratios, unemployment rates, general demographic data, political affiliation of the governor, education expenditures, poverty levels, and NAEP scores by grade and race. These variables come from many sources including the Bureau of Labor Statistics (BLS), Occupational Employment Statistics (OES), Census data, the National Center for Educational Statistics (NCES), the Common Core of Data (CCD), and the Current Population Statistics (CPS). These data sources provide the information necessary to include those factors that vary by state and by time. The recession measure used for this study was calculated by looking at the number of unemployed individuals in each state during each year and dividing that by the number of job vacancies. The appendix of this study describes and summarizes this multitude of controls.

Analysis Models

The basic empirical strategy for this study is to investigate how variation in states' accountability policies impact teachers' work experiences. I begin with a regression model that describes the relationship between teachers' outcomes and accountability stringency:

$$Y_{st} = \beta_0 + \beta_1 AccountabilityStringency + \varepsilon_{st}$$

In this model, Y is the teacher outcome of interest for state s at time t (year) and is regressed against the $AccountabilityStringency$ value for that state and year. ε is the random error in each state and year observation. β_1 is the parameter of interest and describes the relationship between stringency and the given teacher outcome. If β_1 were to equal 1.5 this would indicate that a one percentage point increase in stringency would be associated with a 1.5 unit increase in the outcome, such as the percentage of teachers leaving the profession. This is sometimes referred to as the "naïve estimate" because it would undoubtedly be biased. The naïve regression

does not take into account other trends or policies that impacted or were already impacting

teacher outcomes. For instance, if the number of teachers leaving a state was already declining

prior to the implementation of NCLB, then the estimate from the above regression would be

capturing an already existing trend instead of capturing the impact of NCLB. The following

models represents extensions to the naive model and are utilized to answer the three research

questions of this study.

To capture preexisting trends and to estimate the impact of accountability stringency on

states, a two-way fixed effects approach is utilized that allows for a flexible treatment variable

and multiple time periods spanning pre-NCLB and post-NCLB years. The generalized approach

for continues outcomes is specified below.

$$TeacherOutcomes_{st} = \beta_1 \ln(\bar{F}_{st-}) + X_{st-1}\beta_2 + \theta_s + \delta_t + \varepsilon_{st}$$

This model is used to address the continuous outcomes in this study that are at the state level.

The above model has stringency as a treatment variable and includes state and year fixed effects

for all states and over 20 years. Within the model, $TeacherOutcomes_{st}$, represents teacher

outcomes in each state-year. These outcomes include overall teacher employment numbers, such

as the number of elementary and secondary teachers and the student to teacher ratio. The

treatment variable $\ln(\bar{F}_{st-1})$ reflects a measure of the implementation dosage of NCLB for each

state and year, represented by the log of the simulated AYP failure rates. The simulated AYP

failure rates are calculated using a fixed sample of schools and running each state and year's

rules against this sample providing the percentage of schools failing for that state and year. This

treatment variable provides the advantage of reflecting variation in state accountability policies

while being independent of the population characteristics of the schools and students within the

state. Utilizing this treatment variable aids in reducing the possibility of omitted variable bias,

however, it does not completely address it. The X_{st-1} portion of the equation represents a vector of lagged time varying covariates that were determined prior to the NAEP outcome scores. The fixed effects portion of this model, $\theta_s + \delta_t$, which keeps constant the average effects of either the group or the time period, greatly aids in dealing with the unobserved heterogeneity that impacts the treatment effect and provides a way to control for the average differences across states and years in the observable and unobservable variables.

The state fixed effects portion of the model helps to account for unobserved factors that vary across states, but that do not vary over time and aid in isolating the impact of those variables that do vary over time. One example of the utility of including state fixed effects is the ability to account for an unobservable clustering of teachers into specific states and the gains that are either positively or negatively relate to this unobservable ability. The use of state fixed effects allows for the accounting of this nonrandom distribution of teachers across the states and helps to reduce the bias caused by not accounting for this ability. Another example of the importance of state fixed effects is if states with a high percentage of highly motivated students or an overall "culture" of learning had more positive teacher outcomes. These difficult to quantify motivation and culture factors would bias the results and would confound the actual impact of stringency. Utilizing state fixed effects aids in eliminating these differences in unobserved variables across states that can affect teacher outcomes and helps to isolate the impact of the variation of accountability stringency.

Year fixed effects are especially important to include because they capture the influence of macro factors that affect all states in the same way. If specific events occur within a certain year, such as a large number of teachers retiring, than we would want to capture that as something that is particular and specific to that year. For instance, if different years that have

higher values of negative publicity for education are more likely to have higher stringency values the year fixed effect would be able to account for this development. Also, if an inordinate amount of negative publicity about the teaching profession happened to occur within a single year, teachers themselves may have felt that publicity in a way that is unobserved yet still impactful. Utilizing year fixed effects helps to account for unobserved events across years that can affect teacher outcomes and, like state fixed effects, helps to isolate the impact of the variation of accountability stringency.

Not all outcomes in this study are measured at the state level. The following equation includes a teacher level indicator (i).

$$TeacherOutcomes_{ist} = \beta_1 \ln(\bar{F}_{st-1}) + X_{st-1}\beta_2 + \theta_s + \delta_t + \varepsilon_{ist}$$

This model essentially extends the previous model, but will be used for outcomes at the individual level, which includes the teacher perception and attitude measures taken from the School and Staffing Survey.

For binary outcomes, the following equation is used.

$$Log(\frac{TeacherOutcomes_{ist}}{1 - TeacherOutcomes_{ist}}) = \beta_1 \ln(\bar{F}_{st-1}) + X_{st-1}\beta_2 + \theta_s + \delta_t$$

This equation is used on outcomes that are measured at the individual teacher or student level and are considered binary. These outcomes include whether or not a teacher will be leaving, staying, or moving and whether or not college students are interested in education. Within this equation, I am measuring how changes in stringency impact the probability of an individual making a specific decision.

When investigating how increased accountability stringency differentially affects teacher outcomes based on school characteristics, teacher characteristics, and previous accountability policy strength, the following models were used.

$$TeacherOutcomes_{ist} = \beta_1 \ln(\bar{F}_{st-1}) + X_{ist-1}\beta_2 + (\beta_3 \ln(\bar{F}_{st-1}) * X_{ist-1}) + \theta_s + \delta_t + \varepsilon_{st}$$

$$Log\left(\frac{TeacherOutcomes_{ist}}{1 - TeacherOutcomes_{ist}}\right) = \beta_1 \ln(\bar{F}_{st-1}) + X_{ist-1}\beta_2 + (\beta_3 \ln(\bar{F}_{st-}) * X_{ist-1}) + \theta_s + \delta_t$$

The above model is similar to the previous models, but adds additional variables at the teacher

and school level and looks at the interaction between these variables and stringency. These

variables include important characteristics such as urbanicity, years of experience, age, and

ethnicity. Similarly, I also include within my analytic models the measure for previous

accountability strength and how it interacts with stringency.

Robustness and Sensitivity Checks

One initial sensitivity check run before conducting the full analysis investigates the

balance of covariates before and after the implementation of NCLB. This allows us to check if

our state, teacher, and student populations remain relatively balanced before and after NCLB.

Appendix E illustrates the balance of covariates, with poverty rate and free and reduced lunch the

only characteristics illustrating a significant change. I correct for these changes within my

analytical models and ensure that I take them into account.

As mentioned earlier one key assumption to check when running this analytical model is

the common trend assumption. The common trend assumption says that in the absence of the

accountability stringency of NCLB, the observed outcomes would have followed a parallel trend

across all the states. In other words, checking the common trend assumption would look at the

difference between the observed teacher outcomes taking into account accountability stringency

and the counterfactual teacher outcomes in the absence of NCLB. The violation of the common

trend assumption occurs if something other than the absence of NCLB changes in one state and

not in the other states at the same time of the absent NCLB implementation. The state-year

dataset for this study that includes many states and years allows for the relaxation of the common trends assumption and presents some nonparallel changes in outcomes between states in the absence of NCLB. A model with controls for state-specific trends looks like:

$$TeacherOutcomes_{st} = \beta_1 \ln(\bar{F}_{st-1}) + \mu_s + \alpha_t + \mu_s\alpha_t + XB_{st-} + \varepsilon_{st}$$

This model supposes that without the implementation of NCLB, the teacher outcomes in state s diverge from common year effects (α) and shows the state-specific linear trend by including the interaction between the year and state effects ($\mu_s\alpha_t$). This assesses whether the average change between the states would be equal without the implementation of NCLB. This model also addresses the functional form problem that can arise when doing a two-way fixed effects model.

Extending the model, I investigate the possibility of lagged treatment effects. The impact of stringency may not be immediate and it may take years of stringent accountability or non-stringent accountability to change teacher outcomes. The model, shown in the appendix, utilizes lagged, instead of leading, treatment indicators to test for the delayed effect of treatment.

To test for the appearance of an Ashenfelter's dip, I modify the model to utilize lead treatment indicators and measure the change in slope for states that are about to implement NCLB. If the lead treatment indicators prove to have a significant impact on states, then we could believe that states started reacting to NCLB before the actual implementation of NCLB. The model involves running both one year of a leading indicator, as well as, three years of leading indicators and checking that their estimates are not significantly different from 0. The equations for this check are included in the appendix.

Another extension to the model explores whether history effects are biasing the estimates from the main model. An example of this is the impact of the recession of 2008 on teachers within the state, which may confound the estimates and may be driving the outcomes. To combat

this, I will utilize a comparison group of private school teachers. The private school teachers will face many of the same history issues, such as the recession of 2008, faced by the teachers, but would not be subject in any way to NCLB. This model will include running a model including equivalent outcomes for private school teachers and comparing them to the results gained from running the full stringency model. This equation is included in the appendix.

RESULTS

Figure 2 illustrates national trends in implementation stringency from 2003 to 2011, with some states showing large amounts of variation in stringency and others maintaining a small window of values. Overall, stringency continues to rise over time, rising from an average of around a 45% failure rate to a 59% failure rate. This rise in stringency will inform how the following results will be interpreted and will provide us with an overall implementation effect. Table 17 in Appendix F shows the states that belong to each of the different strength categories, with the fewest number of states belonging to the low level of previous consequential accountability policy group. Figure 11 in Appendix F illustrates the relationship between previous accountability strength and stringency and shows how stringency increases for each of the different groups (no previous accountability policy, weak, moderate, and strong) over the pre-waiver period. However, it is shown that those states that did not have previous accountability policies or had a low level of consequential accountability policies did adopt policies during NCLB that were, on average, more stringent than states that did not have more stringent previous consequential accountability policies. Starting around 2005 and continuing until 2009, states with higher levels of previous consequential accountability policies had NCLB policies that were 10 to 15 percentage points less than the low or no previous accountability policy states. This gap narrowed slightly in 2010 and 2011, but the high and moderate previous accountability strength level states still had, on average, lower levels of stringency than the no and low level states. This does indicate that a state like Iowa (average stringency of 62.5) that had no previous accountability policy before NCLB, but then ramped up to having an above average stringency score over the pre-waiver period would have a lift even above the normal lift of jumping from a non-consequential to a consequential accountability policy as compared to a

state like North Carolina (average stringency of 39.1) that had high levels of consequential accountability policy and could essentially continue business as usual. States with no previous accountability policy had an additional lift, on average, by having adopted more difficult policies during NCLB. Previous research illustrated an impact of this shock to the system, but this study provides a more granular look at the differential impact of this shock.

Table 5 reports the covariates used in the analysis both before and after the start of NCLB. These covariates illustrate that there is only a slight imbalance across the time frames for poverty rate and the percent of students receiving free and reduced lunch. These covariates are used throughout the following analyses. The proceeding section walks through the main effect of stringency on teacher outcomes, the differential impact of stringency based upon previous accountability policies, and the differential impact of stringency based upon specific teacher and school characteristics.

Teacher Employment

Main effects. Figure 3 illustrates changes in overall teacher employment numbers during the pre-waiver period (2003-2011). It shows a slightly oscillating trend in the number of secondary and elementary school teachers per person, however, no significant trends show up in this figure. The current trend of teachers (up to 2016) discussed earlier does show a decrease in the amount of teachers, so investigating the high-level teacher employment numbers during the pre-waiver period may provide some insight into how stringency could have begun to impact teachers.

Tables 6 and 7 present the results of my main impact analyses investigating current teacher employment outcomes, with overall elementary and secondary teacher results in Table 6 and a look at the impact of stringency on teachers leaving, staying and moving in Table 7. Each

column in both tables provides estimates of the coefficient on simulated failure rates from a sequence of models. The base model includes state and year fixed effects. The subsequent models add covariates, such as student and state characteristics, state expenditures, and a final model that includes a measure for recession. The results suggest that more stringent NCLB implementation did not have a significant relationship with the overall employment numbers of teachers per one thousand individuals in a state. Further analysis of the impact of stringency will utilize our full model with a measure for recession, as it captures many of the important factors involved in decisions made by teachers. If significance was found in table 6, for example, using the results from the model in column 5, we would say a one percent change in state accountability stringency increases the number of elementary school teachers per thousand by about 0.0035 teachers per thousand, when controlling for various state and year covariates. Taking this further, we calculate the average implementation effect by taking the average change in stringency (34%) and multiplying it by our coefficient (.0035). In the case of model 5 for elementary school teachers, if significance was found, we would say that the average implementation effect over the 9 years of the study was a paltry .10 elementary school teachers per thousand. Overall, for elementary school teachers, secondary school teachers, and student to teacher ratio, no significant gains or losses could be attributed to increases in accountability stringency.

Table 4 provides a summary of the sample utilized to look at the teachers staying, leaving, and moving from the Teacher Follow-Up Survey. This table illustrates a balanced sample over the years of the analyses, with no significant difference in sample makeup over the years. Figure 4 illustrates an increase of teachers leaving the profession from 1990 to 2012. However, during the pre-waiver period, the trend is less clear, with the actual percentage of

teachers leaving the profession decreasing since 2004-2005. Table 7 is set up similarly to Table 6 with a main impact analysis of stringency on whether or not a teacher will stay, leave, or move within the profession. Each column provides estimates across various models with each coefficient representing a change in the Odds Ratio of an event occurring for a one percent change in state accountability stringency. The model from column 5, which represents a model with state and student characteristics, as well as, a control for recession, illustrates a modestly significant relationship between stringency and teachers moving. No significant results were found for the relationship between stringency and teachers staying or leaving their position. The results indicate a one percent increase in stringency (from baseline) is associated with an increased likelihood of a teacher moving by .3% (Odds Ratio = 1.003, p < .1). Taking this further, the implementation effect of stringency during the pre-waiver period calculates to an odds ratio of 1.095. An odds ratio of 1.095 indicates that stringency increasing by 34% increases the likelihood of a teacher moving by 1.095 times or by 9.5%. Further analysis of why teachers left and its relationship to stringency showed no significant relationship.

Pre-NCLB accountability strength. Taking the analysis further, I look at the previous outcomes with the addition of pre-NCLB accountability strength. Table 8 looks at the effect of strength, stringency, and the interaction between these two on teacher employment numbers. These results were obtained using the recession model and looking at how strong, if any, of an accountability policy a state had before the implementation of NCLB. Table 8 indicates that previous strength had a significant association with teachers' employment decisions in regards to the number of elementary teachers per 1000 (.522, p<.05), the decision to stay in the position (OR = 1.264, p<.01), and leave the position (OR = .865, p<.05). Accounting for strength in the model, stringency also had a significant effect on student to teacher ratios (-2.65, p<.05) and the

employment of elementary school teachers (1.809, p<.1). However, the elementary school teachers per 1000 was the only outcome that illustrated a significant interaction effect between stringency and strength (-1.01, p<.05). This finding can be visualized by looking at the differences in slopes for Figure 5 illustrating that as stringency increases the predicted number of elementary school teachers per 1000 increases only for those states that did not have previous consequential accountability policies. This same distinction is not as defined when looking at the secondary school teachers per 1000. Figure 6 looks at the interaction between strength and stringency for the teacher movement outcomes. This set of graphs indicates that states with either no or low levels of accountability policy before NCLB showed a predicted probability of decrease in teachers moving or leaving and an increase in those staying, while those with high levels of accountability policy before NCLB showed the inverse.

Teacher and school characteristics. Table 9 and Figure 7 shows the results of the interactions between stringency and various teacher and school characteristics to see if stringency differentially impacts based upon these characteristics. These results indicate that the relationship between teacher movement and stringency differed based on characteristics like age, total experience and gender. Among teachers who were between the ages of 25 and 36, increased stringency was negatively related to moving schools and staying at schools, but positively related to leaving the profession all together. In other words, younger teachers were more likely to leave the position than stay within the profession. Differentially, however, looking at the interaction between age and the impact of stringency, we only see significant interactions between age and stringency for the 26-35 year old age group and the 36-50 year old age group. The 26-35 year old age group (OR = .674, p<.05) showed that as stringency increased they were less likely to move,

while the 36-50 year old age group (OR = 1.349, p<.1) showed an indication that they were more likely to move as stringency increased and less likely to leave as stringency increased.

Among female teachers, increased stringency was negatively related to moving out of schools and leaving the profession (OR = .827, p<.1; OR = .694, p<.01), while they have an overall positive association for staying within the profession as compared to male teachers (OR = 1.359, p<.01). However, stringency had a highly significant interaction with female teachers, showing a significant increase in the odds of a female teacher leaving the profession as stringency increases (OR = 1.790, p<.01). This trend is shown in the upper left hand corner of Figure 7 with the slope for female teachers being much steeper than that of male teachers. Finally, no interaction effect between stringency and race or urbanicity was found, indicating the increasing stringency did not differentially impact these separate groups of teachers. This result can be seen when comparing the slopes across figure 7 for these characteristics, as they are not significantly different as stringency increases.

I also examined whether the level of administrative support received by the teacher was associated with teacher's decisions to stay or leave the profession. Administrative support acts as a proxy for principal leadership and illustrates a significant effect on whether teachers stay in the profession, move within the profession, or leave the profession altogether. Running the analysis indicated that administrative support is associated with lowered odds of leaving or moving within the profession.

Teacher Perceptions and Attitudes

Main effects. Although teacher movement appears to have not been highly impacted by changes in stringency, the underlying attitudes and perceptions of teachers about teaching may have been. Figure 8 looks at the changes in time for various measures regarding teachers'

perceptions and attitudes towards their profession during the pre-waiver period. The change over time for these measures was quite variable and for each year the range of values was widespread. Table 10 presents the results of the regression analysis investigating the impact of stringency on the perception measures gathered from the School and Staffing Survey. The results from this analysis indicate that, overall, teacher perceptions and attitudes were not impacted by changing stringency. The lone exception is seen when looking at a measure for salary satisfaction that indicates that stringency had a highly significant negative impact on the raw score on the salary satisfaction scale (-.0011, p < .01). This result indicates that as stringency increases, teachers became less and less satisfied with their current salary, with a one percent increase in stringency being associated with a .0011 point decrease in salary satisfaction. Taking it further, the implementation effect would indicate a .0338 point decrease in salary satisfaction. In terms of the salary satisfaction measure, this would take someone from say a 3.20 out of 4 on the salary satisfaction scale to a 3.16. The practical magnitude of this change is not large, but it does provide some indication that teachers are responding to NCLB.

Pre-NCLB accountability strength. The previous results on the perception analyses indicates that very little impact was found with increasing stringency. However, when accounting for the strength of pre-NCLB accountability policies, as seen in Table 11 and Figure 9, it is shown that pre-NCLB accountability strength has a significant impact and that stringency and strength have a significant interaction for several teacher perception measures. Table 11 illustrates that the differently levels of pre-NCLB accountability policy had a significant effect of control in the classroom and colleague support, with increasing strength being associated with a decrease in control by .3 (p<.05) for each change in strength and an increase in colleague support by .24 (p<.01) for each change in strength. A significant interaction between strength and

stringency is found when looking at their impact on salary satisfaction, administrative support, and colleague support, with the effect of increasing strength providing a stronger negative association between stringency and salary satisfaction (-1.04, p<.01) and colleague support (-.053, p<.05) and the inverse occurring for administrative support (.039, p<.1). As an example, the effect of stringency on salary satisfaction for states is .02, where for each unit increase in stringency it is expected that salary satisfaction would increase by .02 points. This occurs because the effect of strength and the interaction between strength and stringency would be 0. If we investigate the impact of a high pre-NCLB accountability strength (4) than we would multiply the strength effect (.174) by 4, add the effect of stringency (.02) multiplied by the stringency value, and then take the stringency value multiply it by the strength level and then multiple that by the interaction effect (-.104). Since the interaction effect is negative, increasing the strength of the pre-accountability policy sharpens the rate at which salary satisfaction decreases with increasing stringency. These results bear out when looking at Figure 9, specifically with salary satisfaction where the slope for those states with no prior accountability policy illustrates a positive association between salary satisfaction and stringency and the states with a high level of prior accountability stringency show a decline in salary satisfaction as stringency increases.

Teacher and school characteristics. Interacting teacher and school characteristics with stringency as in Table 12 and Figure 10, starts to elucidate certain differential impacts. Overall, teacher and school characteristics had a significant impact on how teachers felt about their position and about the various outcomes, such as control in the classroom, salary satisfaction, administrative support, and colleague support. Colleague support and administrative support both showed similar patterns in how teacher and school characteristics interacted with stringency, as

gender, age and experience all had significant effects on the impact of stringency. As teachers grew older and gained more experience it can be seen that increasing stringency had a stronger association (-.004, p<.01) with decreasing amounts of perceived administrative and colleague support, as well as control in the classroom as compared to younger, less experienced teachers. Female teachers also had a strong association between increasing stringency and a decreasing amount of support from administration (-.078, p<05) and colleagues (-.100, p<.01). This significant difference appears to be driven more by the idea that male teachers reported a higher amount of support as stringency increased, while the slope for female teachers generally remained flat. White and rural teachers grew significantly less satisfied with their salary as stringency increased (-.185 & -.196, p<.01), while urban teachers (.365, p<.01) had the opposite interaction in regards to their satisfaction with salary increasing as stringency increased. Echoing earlier findings, stringency remained significant for many of these subgroups when investigating salary satisfaction.

Potential New Teacher Outcomes

Table 13 presents the results of the analysis on possible new teachers, with the percentage of freshman choosing education majors, percentage of freshman planning on going into an education career, and the percentage of seniors planning on going into an education career. Figure 6 illustrates that over time all three of these percentages drop quite precipitously around 2008. Since this does coincide with the beginning of the recession, we utilize a model that has a measure for recession. This model, shown in Table 13, illustrates that freshmen, when facing increased accountability stringency, choose to forgo a major and/or career in Education. Looking at model 5 in Table 13 we see that the odds of a freshman choosing to go into an Education Major instead of another major decreases by .6% (OR = .994, p < .05) and the choice to go into

an education career decreases by .56% (OR = .994, p < .1) for every one percent increase in stringency. Over the course of the pre-waiver period, the odds of a freshman choosing an education major (OR = .811) or an education career (OR = .835) decreased by 19% and 17% respectively. The same result was not found for seniors' choice of majors or career.

Pre-NCLB accountability strength. Furthermore, choices to go into education as a major or career also show a strong relationship between high or moderate pre-NCLB accountability strength and the impact of stringency. Table 14 and Figure 11 illustrate an initial positive effect on freshman and senior choices towards education, however, when interacted with stringency, it is shown that as strength increases, the impact of increasing stringency has a strong negative association. States with moderate or high pre-accountability strength illustrated a sharp decline in freshmen who were interested in an education major (p<.1) as stringency increased, with stringency decreasing by .30 each time that the pre-accountability strength increases from one level to the next. Similar findings were found for both freshman and senior career choice.

Sensitivity and robustness checks. One threat to validity of the study was the possibility of a response to the recession and housing crisis from 2007 to 2009. This study decided to account for this by including a variable that measured the strength of recession across the states over time. This measure was utilized in all of the regression models run throughout the analyses.

Another threat is the possibility of states enacting changes in stringency as a result of changes in teacher behavior and the possibility of changes in stringency not taking effect until much further down the line. To check for these possibilities I ran models with leading and lagging stringency indicators, looking both forward one year and back a couple of years to see if these leading and lagged treatment variables had an association with outcomes. Table 15 illustrates that there was no significant impact of any of the lagging or leading variables, except

for control in the classroom (p<.1) when looking at stringency + 3 years. This may have occurred by chance, as none of the other key outcomes had a similar finding. Another check I put in place was to look at non-equivalent outcomes, such as things like student problems and family problems which should not be related to stringency in the same way as outcomes like teacher movement and control in the classroom would be. Interestingly, it appears that when looking at a year prior, student and family problems had a strong relationship to stringency. This indicates the possibility that changes in stringency could have been enacted due to student and family behaviors.

Another check conducted by this study was to look at how increasing stringency impacted a group that was equivalent to public school teachers, but that were not under the same requirements. To do so, I investigated private school teachers in the SASS and TFS datasets. Using the full regression model containing a measure for recession, no significant impact of stringency were found on private school teachers across several key outcomes. Table 16 illustrates this point, showing that private school teachers were not significantly impacted by NCLB when controlling for other statewide characteristics that would normally effect both private and public school teachers.

DISCUSSION

The predominant focus of prior research on the impact of NCLB has been on the instantaneous effects of NCLB on students and student performance. Previous work by individuals, such as Mark Schneider (2011), have started to question if those immediate effects of new accountability policy are sustainable. This study shifts focus from students to teachers, with specific focus on how increasing implementation stringency impacts teachers. It also shifts from the instantaneous approach, to a dosage approach by investigating how increasing and variable accountability pressure on schools impacts teachers, while also retaining the idea that immediate effects could have been felt by the implementation of a new policy.

Research Question 1

The first focus of this study was to investigate the impact of stringency on teacher employment outcomes. The previous research on NCLB and teacher employment was mixed, as were the results from this study. I found that current teachers in states with increasing stringency were more likely to consider moving from their current school, although it is unclear if this means out of the state entirely. I did not find any significant impact of stringency on whether teachers leave or stay within the profession, however, I did find that during the NCLB pre-waiver period, potential new teachers in states with higher or increasing stringency were less likely to pursue a career in education. Increased implementation stringency decreased the odds of a freshman choosing an education major or career by 17% and 13% over the study's timeframe. No significant results were found for seniors, suggesting that perhaps freshman were more recently embedded in the consequential accountability framework and also, especially in the later years, would have been under some form of consequential accountability system for most of their educational career. This idea is echoed when looking at the interaction between pre-NCLB

accountability strength and stringency as states with higher pre-NCLB accountability strength indicated a decrease in the likelihood of students pursuing education.

One of the key findings from above was the impact of increasing stringency on college freshman. For the 2017-2018 school year, there was a large teacher shortage (Sutcher, Darling-Hammond, & Carver-Thomas, 2016) and this finding suggests that part of the reason for this shortage may be because of a lack of interest in teaching from the upcoming generation during the pre-waiver period. Follow up research should attempt to look into how to be best attract college students into teaching and also to establish a link between freshman interest in education and the lack of teachers moving through the pipeline.

Research Question 2

The second focus of this study was to look at the impact of stringency on teacher perceptions and attitudes towards their profession. This study was unable to echo the findings of Grissom, Nicholson-Crotty, and Harrington (2014) and other studies (Hamilton et al., 2007; Murnane & Papay, 2010) that found control in the classroom and administrative support improved overall under NCLB. Only minor improvements in the control in the classroom are found with increasing stringency for rural and younger teachers, with negative associations being found for older, more experienced teachers. The only consistent association that arose was the negative impact of increasing stringency on teacher's satisfaction with their salaries. This impacted many of the subgroups within the sample and on the overall sample of teachers. This seems to imply that increasing pressure may cause teachers to have push harder and work longer, leading to a justifiable desire to be compensated appropriately.

Research Question 3

The final focus of this study was to look at how the above outcomes were differentially impacted by different state, school, and teacher characteristics. This study did find consistent differential impacts of stringency across the outcomes. The study found that female teachers were more likely to leave the profession with increasing stringency and that they also found that they did not find an increase in the support from their colleagues or administration as stringency increased. This relationship between support and female teachers leaving could be further explored to investigate a causal connection. The results also pointed out several interesting connections between younger teachers and stringency. Not only did younger teachers feel less of an impact than their older counterparts in regards to moving within the profession, but they also felt a differential increase in control in the classroom and support from colleagues and administration. Much of these results were not found within the overall sample, but when interactions across characteristics occurred, certain patterns emerged.

Pre-NCLB Strength

One final focus of this study was to investigate how pre-NCLB accountability strength differentially impacted teachers and states. There was consistent differential impacts across the strength levels, illustrating that states with tougher pre-NCLB accountability policies were more highly negatively impacted by increased stringency. The findings from this study provide credence to the thought that accountability policy can only reach so far and that increasing stringency, particularly in those states that had previous accountability policy, start to fatigue those that are under the policy. Teacher movement, perceptions, overall employment, and future teacher interest all indicate that there is differential impact based upon how strong the previous accountability policy was for the states. Not only do students, as in Schneider's analysis of Texas

72

(2011), feel stagnation from ongoing accountability policy, but teachers also appear to react to ongoing policy in a similar way. Dee and Jacob (2011) illustrated that the shock of new consequential accountability policy had some positive impacts on teachers and students, however, this study appears to illustrate that on top of a new consequential accountability policy, increasing stringency has a negative impact on those states that did not feel that shock. Even with those states that had no previous consequential accountability policy having higher overall levels of stringency, they still felt some positive impacts of consequential accountability, which was not felt by those states with previous accountability policies. This is an important finding and bears more exploration into how states adopt and implement accountability policies.

CONCLUSION

This study provides a look at the impact of the differential implementation of accountability policies on teachers. Utilizing a measure for stringency, this study was able to assess the impact of variations in accountability policies on teacher movements and teacher perceptions and attitudes. Although the results of the main analysis showed a minimal impact of stringency on teacher outcomes, this study does illustrate that stringency differentially effects teachers based off of gender, age, experience, and on the strength of their state's previous accountability policy. This study provides some indication as to why there are current teacher shortages and provides credence to the idea that teachers may start to wear down from extended periods of high consequential accountability policy. It is possible that with the continuation of consequential accountability to the present day, even those states showing gains from increased stringency during No Child Left Behind may no longer be experiencing the benefits of consequential accountability policies. Further work building off of this study could help to flesh out the teacher experience during periods of consequential accountability and how policy- and decision-makers could help to retain teachers and improve the teacher experience while maintaining high academic standards.

References

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 54–76.

Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators responses to high-stakes testing. *Education Policy Analysis Archives*, *18*, 14. doi: 10.14507/epaa.v18n14.2010

Blinder, A. (2015, April). Atlanta educators convicted in school cheating scandal. *The New York Times*, p. 12A.

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

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review*, *95*(2), 166–171.

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. doi: 10.1177/0013124510362340

Carey, K. (2007). *The Pangloss index: How states game the No Child Left Behind Act.* Washington, DC: Education Sector.

Center on Education Policy (2005). *From the capital to the classroom: Year 3 of the No Child Left Behind Act*. Washington, DC.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, *24*(4), 377–392. doi: 10.1016/j.econedurev.2004.06.008

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources, 41*(4), 778-820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy*, *6*(3), 399–438.

Clotfelter, C., Ladd, H., Vigdor, J., & Aliaga-Diaz, R. (2004). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teacher? *Journal of Policy Analysis and Management,* 23(2), 251-271.

Cullen, J.B. & Reback, R. (2006). Tinkering toward accolades: school gaming under a performance accountability system. *Advances in Applied Microeconomics*, 14, 1–34.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.

Davidson, E., Reback, R., Rockoff, J.E., & Schwartz, H.L. (2013). Fifty ways to leave a child behind: idiosyncrasies and discrepancies in states' implementation of NCLB. *National Bureau of Economic Research (NBER) Working Paper 18988.*

De Mello, V.B. (2011). Mapping state proficiency standards onto NAEP scales: variation and change in state standards for reading and mathematics, 2005-2009. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.

De Mello, V.B. (2015). Mapping state proficiency standards onto NAEP Scales: results from the 2013 naep reading and mathematics assessments. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.

Dee, T. & Jacob, B. (2010). The impact of No Child Left Behind on students, teachers, and schools. *Brookings Papers on Economic Activity*, 149-207.

Dee, T. & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management,* 30(3), 418–46.

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

Evans, W. N., Schwab, R. M., & Wagner, K. L. (2019). The Great Recession and public education. *Education Finance and Policy*, *14*(2), 298–326. doi: 10.1162/edfp_a_00245

Feng, L. (2010). Hire today, gone tomorrow: new teacher classroom assignments and teacher mobility. *Education Finance and Policy*, 5(3), 278–316.

Feng, L. & Sass, T.R. (2008). Teacher quality and teacher mobility, working paper, Florida State University.

Feng, L., Figlio, D.N., & Sass, T.R. (2010). School accountability and teacher mobility. National Bureau of Economic Research (NBER) Working Paper 16070.

Figlio, D.N. (2006). Testing, crime, and punishment. *Journal of Public Economics,* 90, 837–51.

Figlio, D.N. & Getzler, L.S. (2006). Accountability, ability, and disability: gaming the system? *Advances in Applied Microeconomics*, 14, 35–49.

Figlio, D. N. & Kenny, L. W. (2009). Public sector performance measurement and stakeholder support. *Journal of Public Economics*, *93*(9-10), 1069–1077.

Figlio, D. N. & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin & L.

    Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 383-421).

    North-Holland, The Netherlands: Elsevier.

Figlio, D.N. & Rouse, C.E. (2006). Do accountability and voucher threats improve low-

    performing schools? *Journal of Public Economics,* 90(1–2), 239–55.

Fuller, S. C. & Ladd, H. F. (2013). School-based accountability and the distribution of teacher

    quality across grades in elementary school. *Education Finance and Policy*, *8*(4), 528–

    559. doi: 10.1162/edfp_a_00112

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No

    Child Left Behind on teachers' work environments and job attitudes. *Educational*

    Evaluation and Policy Analysis, *36*(4), 417–436.

Halaby, C., & Weakliem, D. (1989). Worker control and attachment to the firm. *American*

    Journal of Sociology, 95(3), 549-591.

Haladyna, T., Haas, N., & Allison, J. (1998). Continuing tensions in standardized

    testing. *Childhood Education*, *74*(5), 262–273. doi: 10.1080/00094056.1998.10521950

Hamilton, L., Stecher, B., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J., et al. (2007).

    Standards-based accountability under No Child Left Behind: Experiences of teachers and

    administrators in three states. Santa Monica, CA: Rand Corporation.

Hamilton, L., Stecher, B., & Yuan, K. (2008). Standards-based reform in the United States:

    history, research, and future directions, Washington, DC: Center on Education Policy.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal*

    of Human Resources, *XXXIX*(2), 326–354. doi: 10.3368/jhr.xxxix.2.326

Hanushek, E.A., & Raymond, M.E. (2005). Does school accountability lead to improved student

performance? *Journal of Policy Analysis and Management,* 24(2), 297–327.

Hanushek, E., & Rivkin, S. (2010). The quality and distribution of teachers under the No Child

Left Behind Act. *The Journal of Economic Perspectives, 24*(3), 133-150.

Harrison-Jones, L. (2007). No Child Left Behind and implications for black students.

The Journal of Negro Education, 76 (3), 346.

Haynes, M. (2014). *On the path to equity: Improving the effectiveness of beginning teachers.*

Washington, DC: Alliance for Excellent Education. Retrieved from

http://all4ed.org/reports-factsheets/path-toequity/

Heilig, H.J. & Darling-Hammond, L (2008). Accountability Texas style: The progress and

learning of urban minority students in a high-stakes testing context. *Educational*

Evaluation and Policy Analysis, 30, 75-110.

Hom, P.W. & Griffeth, R.W. (1995). Employee turnover. Cincinnati, OH: Southwestern College

Publishing.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational

analysis. *American Educational Research Journal*, *38*(3), 499–534. doi:

10.3102/00028312038003499

Ingersoll, R. M., & May, H. (2011). The minority teacher shortage: fact or fable? *Phi Delta*

Kappan, *93*(1), 62–65. doi: 10.1177/003172171109300111

Ingersoll, R., Merrill, L., & May, H. (2016). Do accountability policies push teachers out?

Education Leadership, 73, 44-49.

Ingersoll, R., Merrill, L., & Stuckey, D. (2014). Seven trends: the transformation of the teaching

> force, updated April 2014. CPRE Report (#RR-80). Philadelphia: Consortium for Policy

> Research in Education, University of Pennsylvania.

Ingersoll, R. M., & Perda, D. (2010). Is the supply of mathematics and science teachers

> sufficient? *American Educational Research Journal*, *47*(3), 563–594.

Insights on Educational Policy, Practice, and Research Number 11, (2000). Tough Love: State

> accountability policies push student achievements.

Jacob, B.A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in

> the Chicago Public Schools. *Journal of Public Economics* 89(5–6), 761–96.

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), 843–77.

Kalleberg, A. L., & Wachter, T. M. V. (2017). The U.S. labor market during and after the great

> recession: continuities and transformations. *RSF: The Russell Sage Foundation Journal*

> *of the Social Sciences*, *3*(3), 1. doi: 10.7758/rsf.2017.3.3.01

Karasek, R., Theorell, T. (1990). *Healthy Work*. Basic Books, New York.

Kelly, S., & Northrop, L. (2015). Early career outcomes for the "Best and the

> Brightest." *American Educational Research Journal*, *52*(4), 624–656.

Klein, A. (2016, March 31). The Every Student Succeeds Act: An ESSA overview. Education

> Week. Retrieved Month Day, Year from http://www.edweek.org/ew/issues/every-student-
> succeeds-act/

Krei, M. S. (1998). Intensifying the barriers. *Urban Education*, *33*(1), 71–94. doi:

> 10.1177/0042085998033001005

Krieg, J.M. (2008). Are students left behind? The distributional effects of No Child Left Behind. *Education, Finance and Policy* 3(2), 250–81.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions. *Educational Evaluation and Policy Analysis*, *33*(2), 235–261. doi: 10.3102/0162373711398128

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. doi: 10.3102/01623737024001037

Linn, R.L. (1999). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44–70. doi: 10.1207/s15327930pje8003_4

Mueller, C. W., & Price, J. L. (1990). Economic, psychological, and sociological determinants of voluntary turnover. *Journal of Behavioral Economics*, *19*(3), 321–335. doi: 10.1016/0090-5720(90)90034-5

Mulvenson, S. W., Stegman, C. E., & Ritter, G. (2005). Test anxiety: A multifaceted study on the perceptions of teachers, principals, counselors, students, and parents. *International Journal of Testing*, 5(1), 37-61.

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. doi: 10.1257/jep.24.3.151

Neal, D. & Schanzenbach, D.W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics,* 92(2), 263–83.

O'Day, J. A., and Smith, M. S. (1993). "Systemic reform and educational opportunity." In S. H. Fuhrman (Ed.), Designing Coherent Education Policy: Improving the System. San Francisco: Jossey-Bass, 250-312.

Price, J.L. (1977) The study of turnover. Iowa State University Press, Ames.

Price, J. L. (1989). The impact of turnover on the organization. *Work and Occupations*, *16*(4), 461–473. doi: 10.1177/0730888489016004005

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics* 92(5–6), 1394–1415.

Reback, R., Rockoff, J., & Schwartz, H.L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind: Dataset. *American Economic Journal: Economic Policy*. http://dx.doi.org/10.1257/pol.6.3.207.

Rice, J.K. (2013). Learning from experience: Evidence on the impact and distribution of teacher experience and the implications for teacher policy. *Education Finance and Policy*, 8(3), 332-348.

Rogasa, D. (2003). The "99% confidence" scam: Utah-style calculations. Stanford University.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36.

Rubin, D. (2011). *The disheartened teacher: Living in the age of standardization, high-stakes assessments, and No Child Left Behind (NCLB)*. Changing English, 18, 407-416. doi.org./10.1080/1358684X.2011.630197

Santoro, D. A. (2011). Good teaching in difficult times: Demoralization in the pursuit of good work. *American Journal of Education*, *118*(1), 1–23. doi: 10.1086/662010

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher

    mobility. *Economics of Education Review*, *26*(2), 145–159. doi:

    10.1016/j.econedurev.2005.08.006

Schneider, M. (2011). The Accountability Plateau, *Thomas B. Fordham Institute*, at

    https://files.eric.ed.gov/fulltext/ED528959.pdf. (January 17, 2020).

Sun, M., Saultz, A., & Ye, Y. (2016). Federal policy and the teacher labor market: exploring the

    effects of NCLB school accountability on teacher turnover. *School Effectiveness and*

    School Improvement, *28*(1), 102–122. doi: 10.1080/09243453.2016.1242506

Sunderman, G., Tracey, A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom*

    realities and No Child Left Behind. Cambridge, MA: Civil Rights Project, Harvard

    University.

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016). *A coming crisis in teaching?*

    Teacher supply, demand, and shortages in the U.S.. Palo Alto, CA: Learning Policy

    Institute.

Wei, X. (2012). Are more stringent NCLB state accountability systems associated with better

    student outcomes? An analysis of NAEP results across states. *Educational Policy*. 26(2),

    268-308.

Wong, M., Cook, T. D., & Steiner, P. M. (2015). Adding design elements to improve time series

    designs: No Child Left Behind as an example of causal pattern-matching. *Journal of*

    Research on Educational Effectiveness, *8*(2), 245–279.

Wong, V. C., Wing, C., Martin, D., & Krishnamachari, A. (2017). did states use implementation

    discretion to reduce the stringency of NCLB? Evidence from a database of state

    regulations. *Educational Researcher*, *47*(1), 9–33.

Appendix A: Tables and Figures
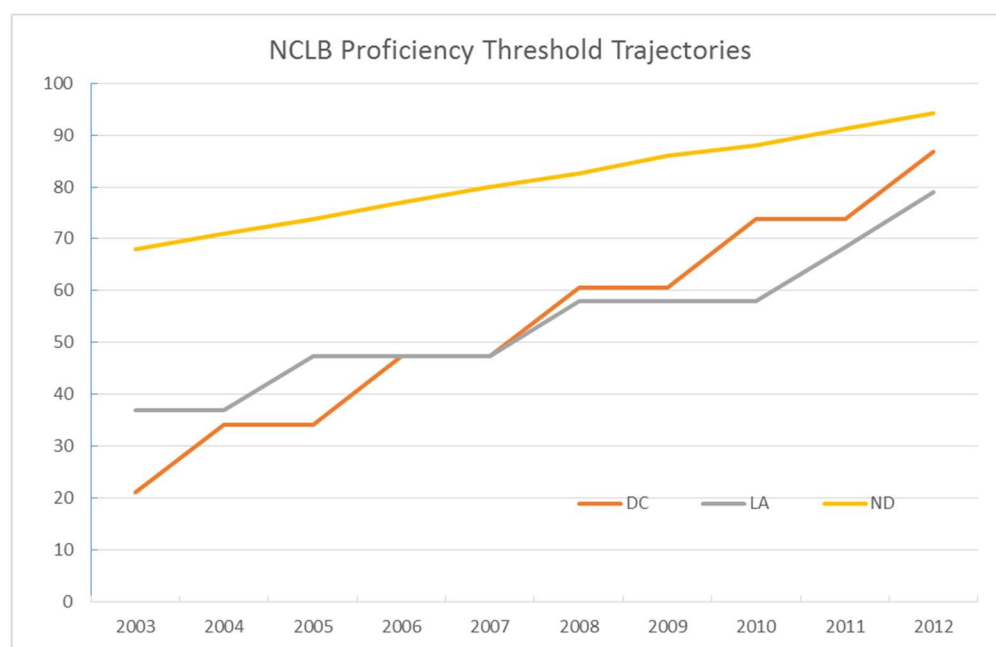
Figure 1: Examples of Proficiency Threshold Trajectories

Table 1: Characteristics of the NAEP Sample

Year	Total Count	White	Black	Hispanic	Asian	FRL
2009	1,658,753	923,693	293,910	310,572	79,453	776,314
		55.7%	17.7%	18.7%	4.8%	46.8%
2007	1,713,181	968,183	308,620	307,419	78,964	753,751
		56.5%	18.0%	17.9%	4.6%	44.0%
2005	1,763,542	986,361	302,276	282,182	74,786	740,411
		55.9%	17.1%	16.0%	4.2%	42.0%

Table 2: Coding Fidelity

	High Coding Fidelity	Partial Coding Fidelity	Low Coding fidelity
		Adjusted AYP Rules	Consequential growth models
	District of Columbia, Georgia, Idaho, Illinois, Indiana, Kansas, Louisiana, Maine, Maryland, Minnesota, Missouri, Montana, Nevada, New Jersey, New Mexico, North Dakota, Oregon, Pennsylvania, South Carolina, South Dakota, Texas, Utah, Virginia, West Virginia, Wisconsin	Alabama, Alaska*, Arizona*, Arkansas*, California, Colorado, Connecticut, Delaware*, Florida*, Hawaii, Iowa*, Kentucky, Massachusetts, Michigan, Mississippi, Nebraska, New Hampshire, New York, North Carolina*, Oklahoma, Rhode Island, Tennessee*, Vermont, Washington, Wyoming	Ohio
N	25	25	1

* In addition to adjusted AYP rules, these states' also included inconsequential growth model rules (N = 8)

Table 3: AYP Calculator Validation Results

	Actual AYP Failure Rates	*Predicted* AYP Failure Rates
California (2004-05)	39%	40%
California (2010-11)	29%	32%
Pennsylvania (2003-04)	14%	14%
Pennsylvania (2007-08)	28%	27%
Texas (2003-04)	17%	16%
Texas (2007-08)	34%	36%
Maryland (2005-06)	23%	30%
Maryland (2009-10)	33%	32%
New York (2009-10)	38%	38%
New York (2010-11)	47%	50%
North Carolina (2009-10)	42%	41%
North Carolina (2010-11)	29%	22%
Ohio (2006-07)	35%	38%
Ohio (2010-11)	40%	55%
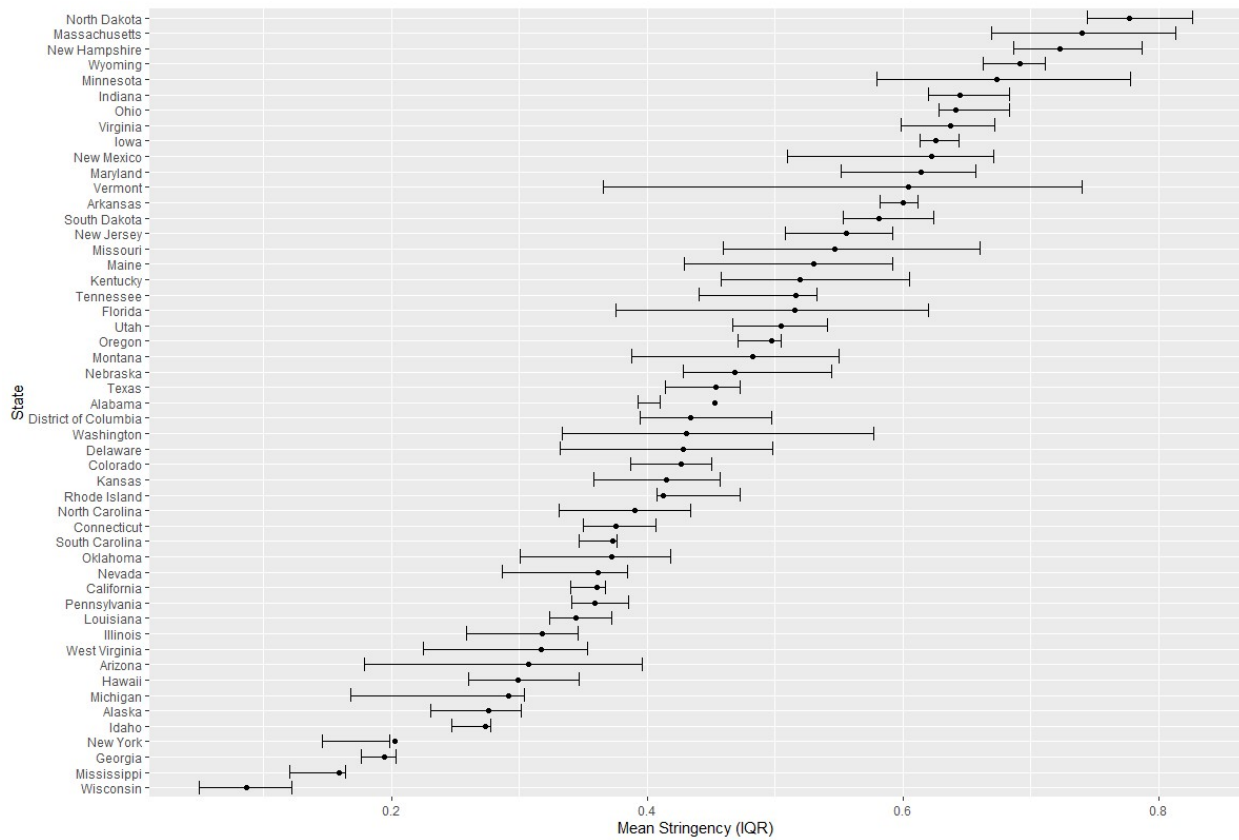
Figure 2: Mean Stringency and Inner Quartile Range in all states from 2003 to 2011

Table 4: SASS/TFS Sample Information

		2004	2008	2012	p value
Female n(%)		16771 (77.1)	14920 (77.8)	15703 (75.3)	0.911
White n(%)		18214 (83.7)	16678 (86.9)	18405 (88.2)	0.982
Age					
	< 25	1280 (5.9)	1289 (6.7)	1436 (6.9)	
	26 - 35	5705 (26.2)	5060 (26.4)	6128 (29.4)	
	36 - 50	8404 (38.6)	7008 (36.5)	7717 (37.0)	0.9995
	51 - 65	6265 (28.8)	5681 (29.6)	5426 (26.0)	
	66+	108 (.50)	148 (.77)	158 (.76)	

Table 5: Covariate Balance - Pre/Post NCLB

	Pre-NCLB		Post-NCLB		
	Mean	Standard Deviation	Mean	Standard Deviation	P
State Population Characteristics					
State Population	5701744.63	6426006.79	5919267.35	6628288.53	0.825
Poverty Rate	11.75	3.22	12.73	3.28	0.043
Unemployment Rate	5.65	1.05	6.25	2.32	0.067
% with Bachelor's Degree	26.2	4.61	26.84	4.42	0.38
% Republican (Governor's Affiliation)	55	50	49	50	0.39
Teacher Characteristics					
% Less than 35 Years Old	16.03	2.55	15.11	3.35	0.063
% White	84.13	16.03	84.95	16.2	0.743
% Black	7.69	12.29	7.07	11.97	0.736
% Hispanic	3.15	5.15	3.63	5.65	0.575
Pupil Teacher Ratio	15.5	2.32	15.14	2.65	0.352
Student Characteristics					
% Black Students	15.72	16.48	15.71	15.91	0.997
% White Students	67.71	19.62	64.24	19.64	0.232
% Hispanic Students	9.89	11.62	12.25	12.62	0.202
% of Students on FRPL	29	11	33	10	0.023

Figure 3: Overall Teacher Employment Numbers

Table 6: Overall Teacher Employment Analysis Results:

	Elementary School Teachers				
	1	2	3	4	5
Simulated Failure Rate (logged)	0.0056	0.0051	0.0043	0.0038	0.0035
	(0.0049)	(0.0051)	(0.0049)	(0.0047)	(0.0042)
Implementation Effect	0.1639	0.1493	0.1258	0.1112	0.1024
	Secondary School Teachers				
	1	2	3	4	5
Simulated Failure Rate (logged)	-0.0030	-0.0042	-0.0037	-0.0035	-0.0031
	(0.0040)	(0.0037)	(0.0036)	(0.0035)	(0.0037)
Implementation Effect	-0.0878	-0.1229	-0.1083	-0.1024	-0.0907
	Student to Teacher Ratio				
	1	2	3	4	5
Simulated Failure Rate (logged)	-0.0054	-0.0051	-0.0052	-0.0055	-0.0067
	(0.0049)	(0.0054)	(0.0050)	(0.0052)	(0.0056)
Implementation Effect	-0.1580	-0.1493	-0.1522	-0.1610	-0.1961
Model Specifications for each column					
State FE	x	x	x	x	x
Year FE	x	x	x	x	x
State Characteristics		x	x	x	x
Student Characteristics			x	x	x
State Expenditures				x	x
Recession					x
N Years	9	9	9	9	9
N States	48	48	48	48	48
N Observations	432	432	432	432	432

Depending on the model specification, the Simulated Failure Rate coefficient is estimated by separate linear regressions of teacher employment on simulated failure rates, state and year fixed effects, time-varying covariates, and a measure for the strength of recession.

The coefficients illustrates the raw change in teacher employment numbers for a one percent change in stringency. For example, in model 5, a one percent change in stringency would be associated with an increase of .0035 elementary school teachers per 1000.

The implementation effect is the difference in teacher employment numbers that is the result of the average percent change in simulated failure rates from 2003 to 2011 (34%). This was calculated by $\beta \times \ln(1.34)$. Standard errors are clustered at the state level.
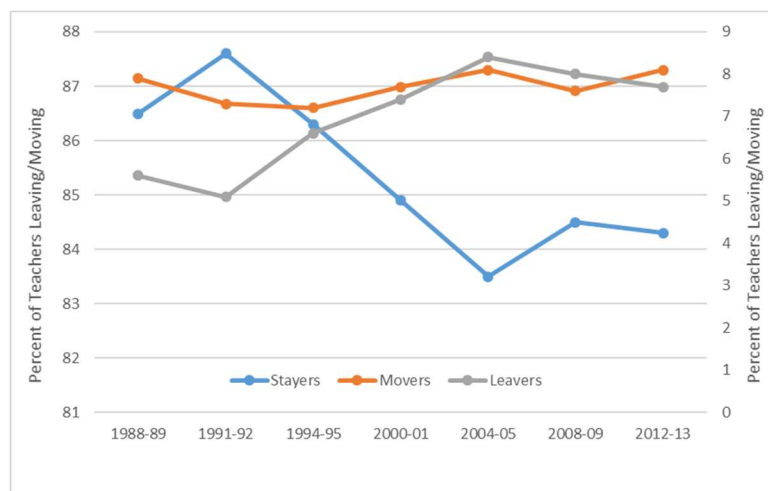
Figure 4: Movers, Stayers, Leavers over time

Table 7: Teacher Follow-up Survey Analysis

	Teachers Leaving				
	1	2	3	4	5
Simulated Failure Rate (logged)	0.9993	0.9996	0.9992	0.9995	0.9999
Implementation Effect	0.9785	0.9871	0.9776	0.9867	0.9980
	Teachers Staying				
	1	2	3	4	5
Simulated Failure Rate (logged)	0.9993	0.9992	0.9987	0.9984	0.9983
Implementation Effect	0.9810	0.9769	0.9638	0.9548	0.9526
	Teachers Moving				
	1	2	3	4	5
Simulated Failure Rate (logged)	1.0019	1.0019	1.0030*	1.0033**	1.0031*
Implementation Effect	1.0581	1.0563	1.0916	1.1027	1.0955
Model Specification					
State FE	x	x	x	x	x
Year FE	x	x	x	x	x
State Characteristics		x	x	x	x
Student Characteristics			x	x	x
State Expenditures				x	x
Recession					x

*p<0.10, ** p<0.05, *** p<.01

Depending on the model specification, the Simulated Failure Rate coefficient is estimated by separate logistic regressions of teacher movement on simulated failure rates, state and year fixed effects, time-varying covariates, and a measure for the strength of recession.

The coefficients are represented as Odds Ratios for logistic regressions in this study. The coefficient represents the change in the likelihood of an event occurring (coefficient - 1) for a one percent increase in stringency. For example, in model 5, a one percent increase in stringency decreases the likelihood of a teacher staying in the profession vs. not staying in the profession by .0017 (.9983 − 1) or .17% (.0017*100).

The implementation effect is the change in the likelihood of teacher movement that is the result of the average percent change in simulated failure rates from 2003 to 2011 (34%). This was calculated by $\beta \times \ln(1.34)$.

Table 8: Employment Outcomes and Pre-NCLB Accountability

	Strength Effect	Stringency Effect	Interaction Effect
Overall Teacher Employment			
Elementary per 1000	0.522**	1.809*	-1.01**
Secondary per 1000	-0.117	0.124	0.038
Student/Teacher Ratio	-0.598	-2.65**	0.577
Teacher Movement			
Mover	1.190	1.102	1.122
Stayer	1.264***	1.092	0.948
Leaver	0.865**	0.898	0.969

*p<.10, **p<.05, ***p<.01

Overall Teacher Employment outcomes were estimated using a linear regression on simulated failure rates, state and year fixed effects, time-varying covariates, a measure for the strength of recession, the strength of pre-NCLB accountability and an interaction between strength of pre-NCLB accountability and stringency.

Teacher Movement outcomes were estimated using a logistic regression with the same model specifications as above.

The coefficients for Overall Teacher Employment represent raw changes in teacher employment numbers and the coefficients for Teacher Movement are illustrated in Odds Ratios.

The strength effect represents the impact of moving from one level of pre-NCLB accountability strength to the next level (i.e.: Weak to Moderate).

The stringency effect represents the impact of a one percent increase in stringency.

The interaction effect is the differential effect of stringency on the various levels of pre-NCLB accountability. For example, as the strength of pre-NCLB accountability increases the negative impact of stringency on Elementary School teachers per 1000 significantly increased (-1.01 per level).
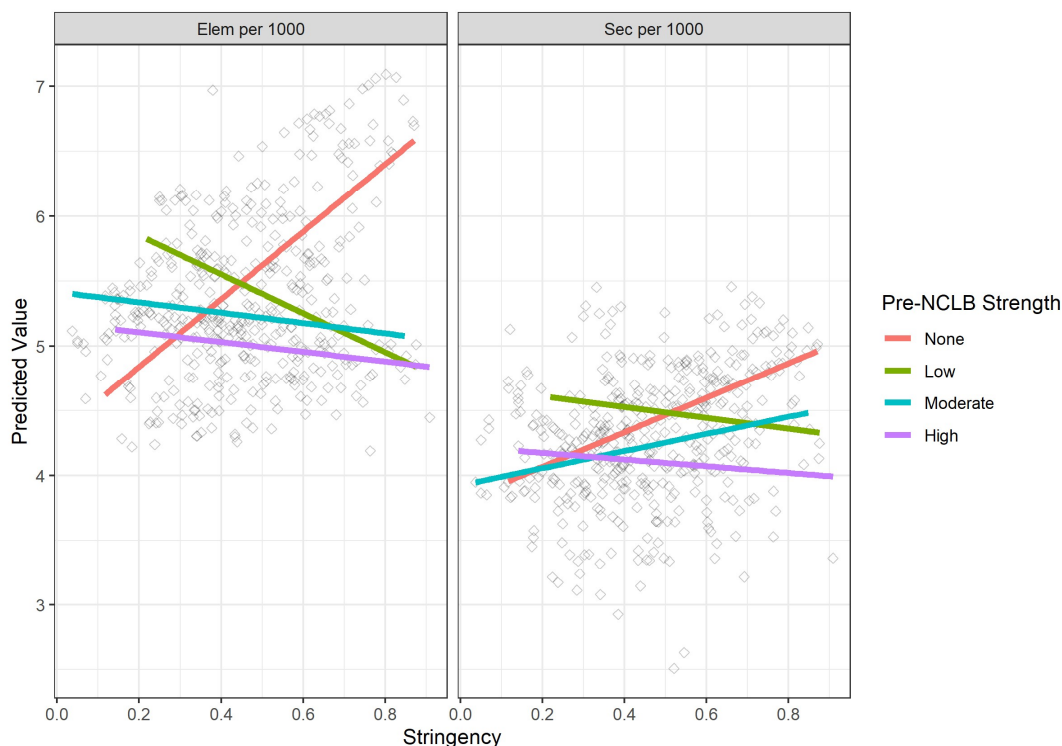
Figure 5: Employment Outcomes and pre-NCLB Accountability Strength
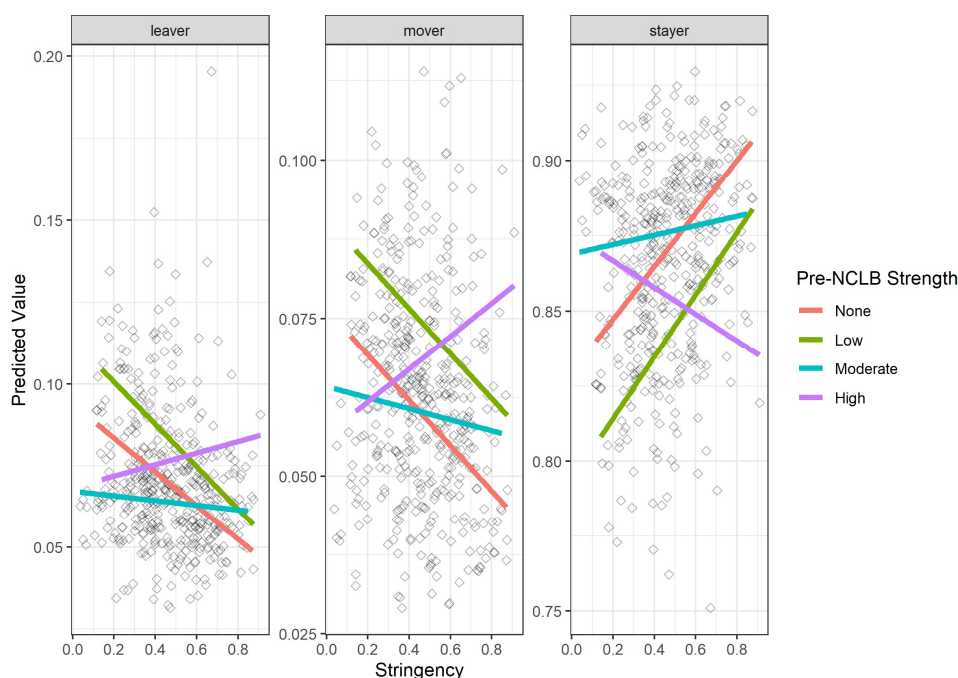
Figure 6: Teacher Movement and pre-NCLB Accountability Strength

Table 9: Teacher Movement and Teacher/School Characteristics

	Mover			Leaver			Stayer		
	Effect	Stringency	Interaction	Effect	Stringency	Interaction	Effect	Stringency	Interaction
White	0.989	1.644**	0.789	0.932	1.106	0.866	1.047	0.721*	1.230
Female	0.827*	1.351	1.019	0.694***	0.641**	1.790***	1.359***	1.087	0.714**
Urban	1.064	1.185	1.051	1.259**	1.094	0.978	0.847**	0.865	0.991
Rural	0.948	1.250	0.877	0.960	1.119	0.903	1.051	0.832	1.135
Age < 25	0.500***	1.328	1.404	1.314*	0.989	1.057	0.581***	0.862	0.780
Age 26-35	2.034***	1.573**	0.674**	0.905	0.950	1.175	0.716***	0.816	1.134
Age 36-50	0.719***	1.231	1.349*	1.939***	1.093	0.658**	1.744***	0.845	0.996
Age 51-65	0.460***	1.365*	0.919	1.833***	0.964	1.125	0.948	0.848	0.993
Total Experience	0.944***	1.306	1.006	1.010**	0.772	1.017**	1.019***	0.975	0.989*

*p<.10, **p<.05, ***p<.01

Teacher Movement outcomes were estimated using a logistic regression on simulated failure rates, state and year fixed effects, time-varying covariates, a measure for the strength of recession, the teacher/school characteristic and an interaction between the teacher/school characteristic and stringency.

The coefficients are represented as Odds Ratios for logistic regressions in this study.

Effect represents the impact of belonging to one group vs. another (i.e.: Female vs. Male).

Stringency represents the impact of a one percent increase in stringency.

Interaction is the differential effect of stringency on the different levels of the covariate. If the coefficient is above 1, than the impact of stringency will be greater for that given characteristic.
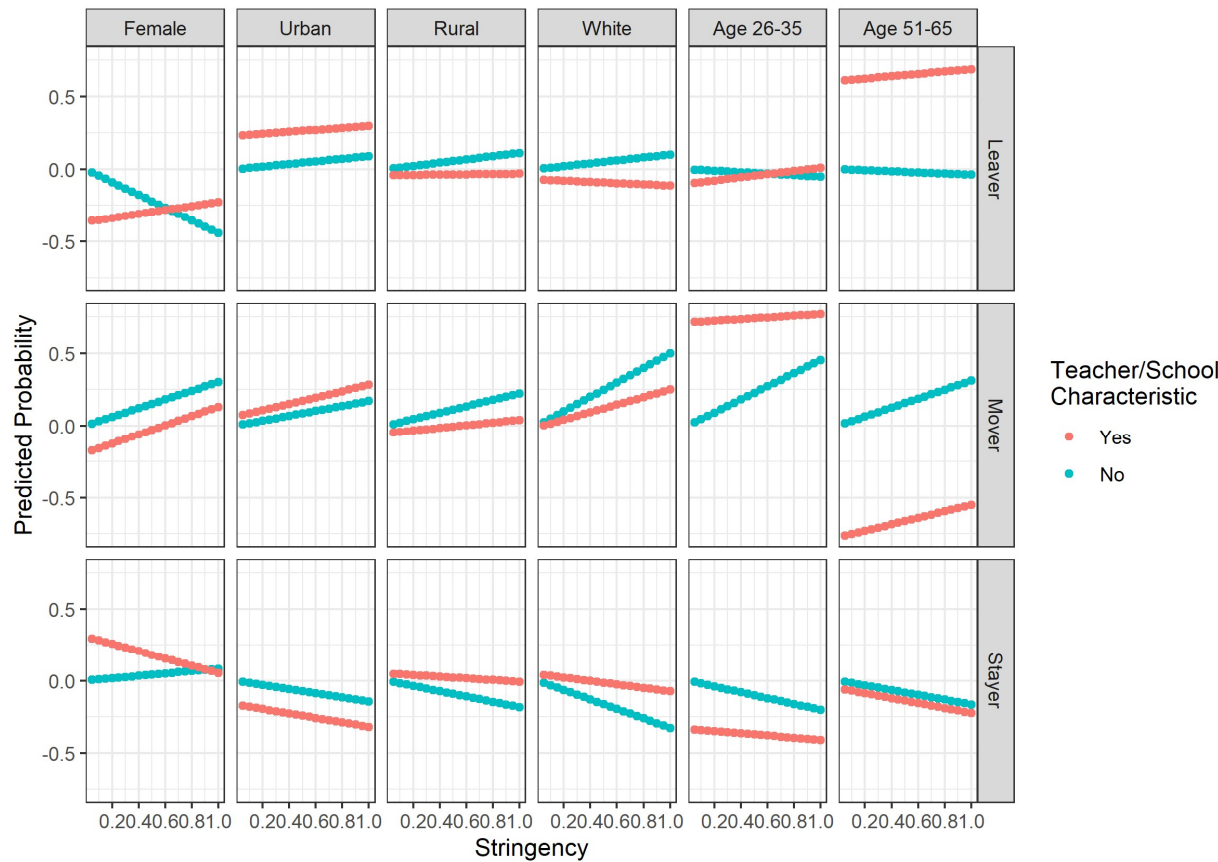
Figure 7: Teacher Movement and Teacher/School Characteristics
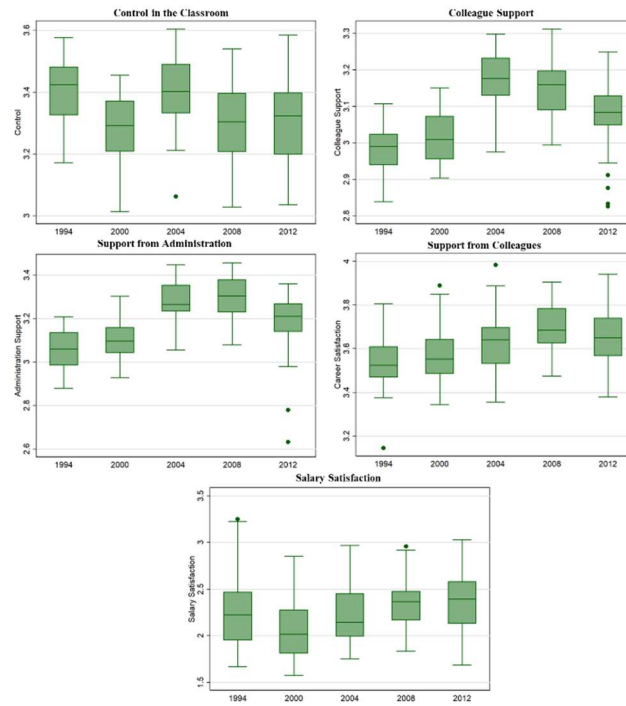
Figure 8: Selection of Teacher Perception outcomes over time

Table 10: Teacher Perceptions and Stringency Analysis

	Control in the Classroom			Admin Support		
	1	2	3	1	2	3
Simulated Failure Rate (logged)	0.0003	0.0002	0.0002	-0.0002	0.0001	0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Implementation Effect	0.0085	0.0071	0.0067	-0.0052	0.0024	0.0059
	Colleague Support			Salary Satisfaction		
	1	2	3	1	2	3
Simulated Failure Rate (logged)	0.0000	0.0002	0.0002	-0.0008**	-0.0011***	-0.0011***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0004)
Implementation Effect	0.0004	0.0047	0.0056	-0.0241	-0.0326	-0.0338
	Career Satisfaction			Job Satisfaction		
	1	2	3	1	2	3
Simulated Failure Rate (logged)	-0.0002	-0.0001	-0.0001	-0.0002	-0.0002	-0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Implementation Effect	-0.0072	-0.0039	-0.0032	-0.0072	-0.0053	-0.0027
	Burnout			Burnout/Job/Career Satisfaction		
	1	2	3	1	2	3
Simulated Failure Rate (logged)	-0.0004	-0.0004	-0.0003	-0.0004	-0.0004	-0.0003
	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)
Implementation Effect	-0.0115	-0.0124	-0.0095	-0.0121	-0.0112	-0.0091
Model Specifications						
State FE	x	x	x	x	x	x
Year FE	x	x	x	x	x	x
State Characteristics	x	x	x	x	x	x
Student Characteristics	x	x	x	x	x	x
State Expenditures		x	x		x	x
Recession			x			x
N Years	3	3	3	3	3	3
N States	49	49	49	49	49	49
N Observations	137331	137331	137331	137331	137331	137331

*p<0.10, ** p<0.05, *** p<.01

The Simulated Failure Rate coefficient is estimated by separate linear regressions of teacher perceptions/attitudes on simulated failure rates, state and year fixed effects, time-varying covariates, and a measure for the strength of recession. (Fixed Effect/State Effect only and State Characteristic Models are available in the appendix)

The coefficient illustrates the raw change in a given outcome/perception score for a one percent change in stringency. In general, these values were rated on a scale of 1 to 4. For example, in model 3, a one percent increase in stringency would be associated with a decrease in salary satisfaction of .0011.

The implementation effect is the difference in perception score that is the result of the average percent change in simulated failure rates from 2003 to 2011 (34%). This was calculated by $\beta \times \ln(1.34)$.

Table 11: Teacher Perceptions and pre-NCLB Accountability Strength

	Strength Effect	Stringency Effect	Interaction Effect
Teacher Perceptions			
Control	-.300**	0.06	-0.021
Job Satisfaction	-0.044	-0.032	0.0008
Salary Satisfaction	0.174	0.02	-.104***
Admin Support	-0.085	-0.044	.039*
Colleague Support	.240***	0.088	-0.053**
Family Problems	0.184	-0.01	0.035

*p<.10, **p<.05, ***p<.01

Teacher Perception outcomes were estimated using a linear regression on simulated failure rates, state and year fixed effects, time-varying covariates, a measure for the strength of recession, the strength of pre-NCLB accountability and an interaction between strength of pre-NCLB accountability and stringency.

The coefficients illustrates the raw change in a given outcome/perception score.

The strength effect represents the impact of moving from one level of pre-NCLB accountability strength to the next level (i.e.: Weak to Moderate).

The stringency effect represents the impact of a one percent increase in stringency.

The interaction effect is the differential effect of stringency on the various levels of pre-NCLB accountability. For example, as the strength of pre-NCLB accountability increases the positive impact of stringency Administrative Support significantly increased (.039 per change in level).
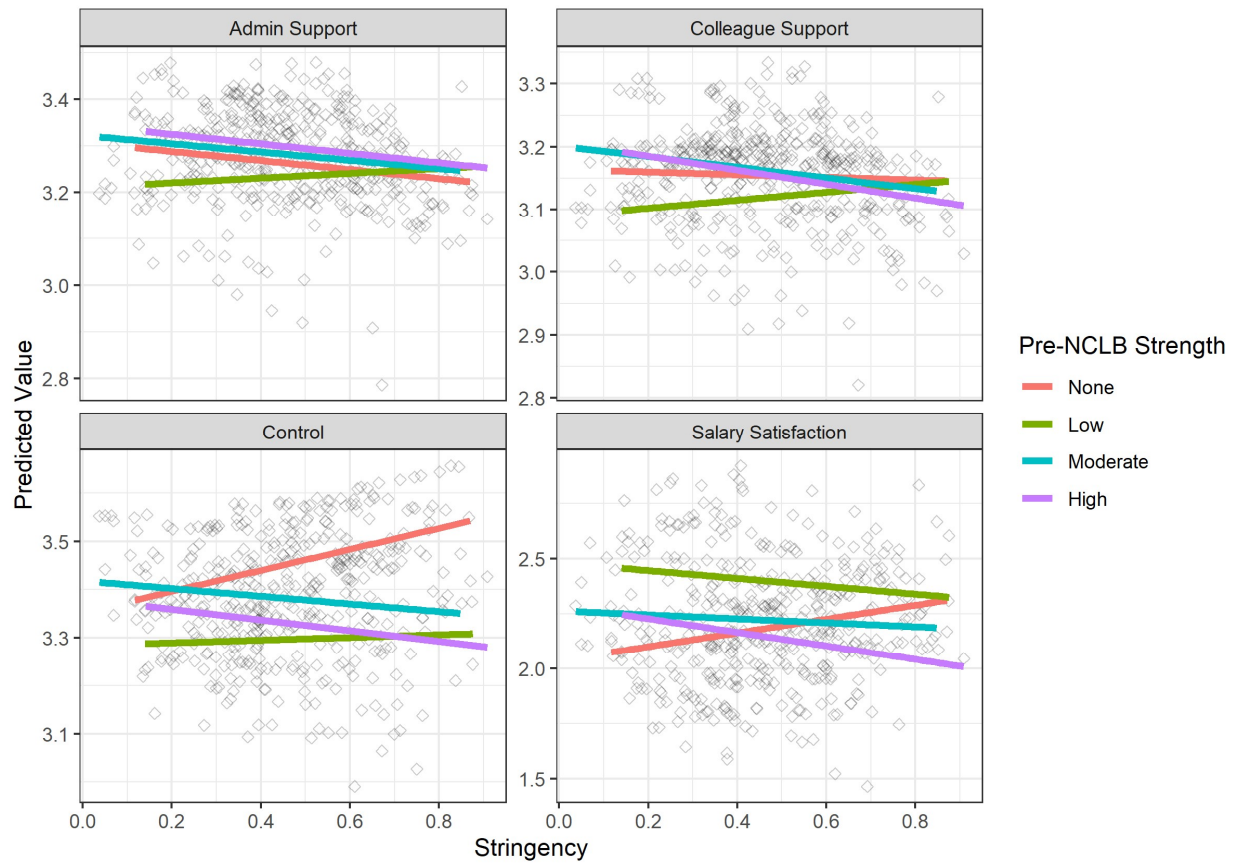
Figure 9: Teacher Perceptions and pre-NCLB Accountability Strength

Table 12 Teacher Perceptions and the Impact of Teacher/School Characteristics:

	Control			Salary Satisfaction		
	Effect	Stringency	Interaction	Effect	Stringency	Interaction
White	.032**	0.011	0.016	.181***	0.044	-.185***
Female	-.058***	0.045	-0.027	-0.039	-.151***	0.047
Urban	-.138***	0.028	-0.002	-.226***	-.216***	.365***
Rural	.110***	0.018	.045*	.141***	-0.05	-.196***
Age < 25	-.088**	0.016	.131***	.190***	-.113***	-0.03
Age 26-35	-.032**	0.009	.051**	-.049**	-.124***	0.03
Age 36-50	.038***	0.046	-.061***	0.008	-.121***	0.014
Age 51-65	0.009	0.027	-0.014	-0.033	-.111**	-0.021
Total Experience	.003***	.075***	-.004***	-0.001	-0.082	-0.002
	Admin Support			Colleaague Support		
	Effect	Stringency	interaction	Effect	Stringency	interaction
White	0.003	0.019	0.001	0.023	0.015	0.006
Female	.035*	.080**	-.078**	.166***	.09**	-.100***
Urban	-0.024	0.01	-0.009	-.055***	0.004	0.054
Rural	-.067***	-0.019	.074**	-.043***	0.005	0.037
Age < 25	.109***	0.025	-0.08	.077***	0.027	-.137**
Age 26-35	-.037**	-0.0002	.074**	-.141***	-0.01	.106***
Age 36-50	-0.019	-0.016	0.011	0.011	0.031	-0.03
Age 51-65	0.025	0.039	-.067**	.100***	0.03	-0.032
Total Experience	0.001	.075**	-.004***	.005***	.071**	-.004***

*p<.10, **p<.05, ***p<.01

Teacher Perception outcomes were estimated using a linear regression on simulated failure rates, state and year fixed effects, time-varying covariates, a measure for the strength of recession, the teacher/school characteristic and an interaction between the teacher/school characteristic and stringency.

The coefficients illustrates the raw change in a given outcome/perception score.

Effect represents the impact of belonging to one group vs. another (i.e.: Female vs. Male).

Stringency represents the impact of a one percent increase in stringency.

Interaction is the differential effect of stringency on the different levels of the covariate. For example, the impact of stringency on Control in the Classroom is .045 times higher in Rural teachers compared to non-Rural teachers..
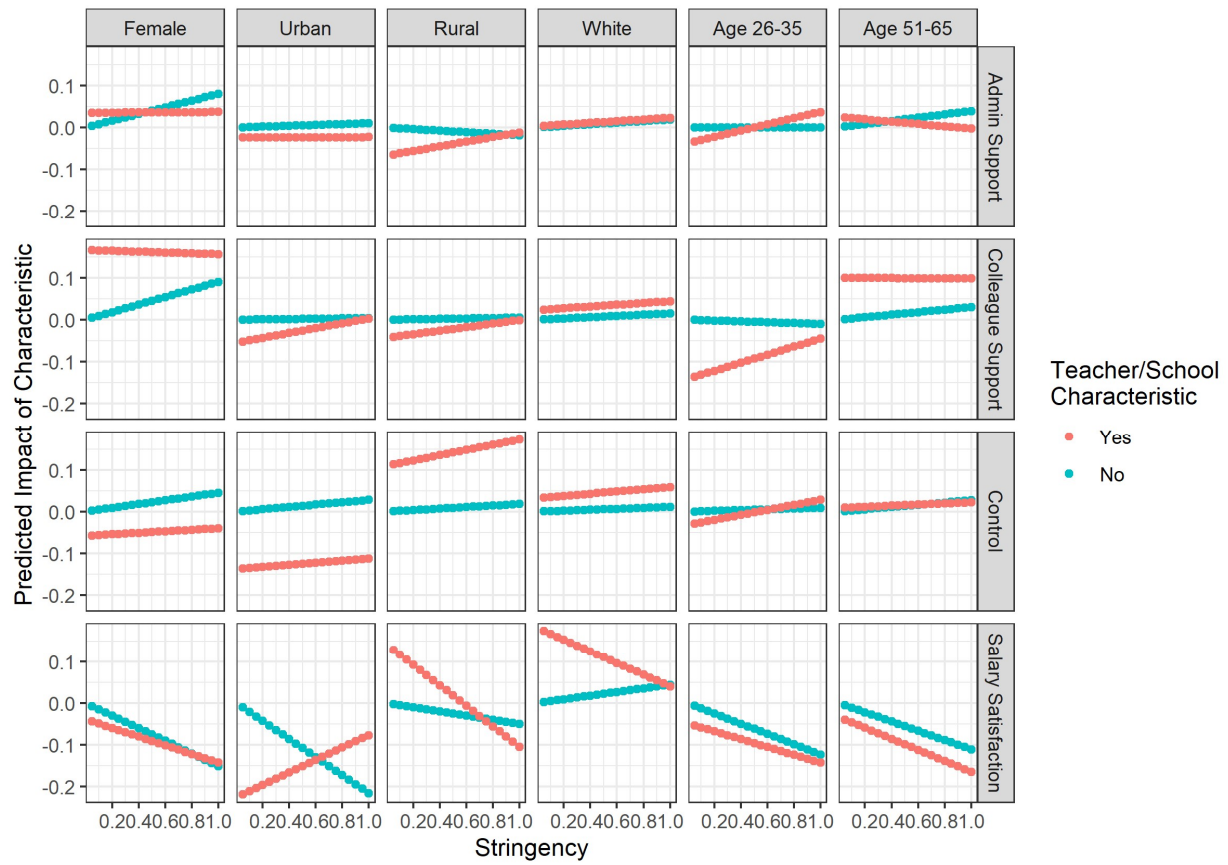
Figure 10: Teacher Perceptions and the Impact of Teacher/School Characteristics:

Table 13: Student Choices and Stringency

	Major Freshman				
	1	2	3	4	5
Simulated Failure Rate (logged)	0.9954	0.9950	0.9950	0.9942	0.9935**
Implementation Effect	0.8623	0.8513	0.8513	0.8296	0.8111
	Career Choice Freshman				
	1	2	3	4	5
Simulated Failure Rate (logged)	0.9962	0.9958	0.9957	0.9948*	0.9944*
Implementation Effect	0.8848	0.8735	0.8707	0.8458*	0.8349*
	Career Choice Seniors				
	1	2	3	4	5
Simulated Failure Rate (logged)	0.9998	0.9995	0.9999	0.9991	0.9988
Implementation Effect	0.9841	0.9512	0.9911	0.9148	0.8869
Model Specifications					
State FE	x	x	x	x	x
Year FE	x	x	x	x	x
State Characteristics		x	x	x	x
Student Characteristics			x	x	x
State Expenditures				x	x
Recession					x
N Years	9	9	9	9	9
N States	49	49	49	49	49
N Observations	121589	121589	121589	121589	121589

*p<0.10, ** p<0.05, *** p<.01

Depending on the model specification, the Simulated Failure Rate coefficient is estimated by separate logistic regressions of student choices on simulated failure rates, state and year fixed effects, time-varying covariates, and a measure for the strength of recession.

The coefficients are represented as Odds Ratios for logistic regressions in this study. The coefficient represents the change in the likelihood of an event occurring (coefficient - 1) for a one percent increase in stringency. For example, in model 5, a one percent increase in stringency decreases the likelihood of a freshman choosing an education major vs. choosing a different major by .006 (.994 − 1) or .6% (.006*100).

The implementation effect is the change in the likelihood of student choices that is the result of the average percent change in simulated failure rates from 2003 to 2011 (34%). This was calculated by $\beta \times \ln(1.34)$

Table 14: Student Choice and pre-NCLB Accountability Strength

	Strength Effect	Stringency	Interaction
Freshman Ed Major	1.162*	0.869	0.741*
Freshman Ed Career	1.150*	0.795	0.748*
Senior Ed Career	1.105*	0.980	0.787**
*p<.10, **p<.05, ***p<.01			

Student choice outcomes were estimated using a logistic regression on simulated failure rates, state and year fixed effects, time-varying covariates, a measure for the strength of recession, the strength of pre-NCLB accountability and an interaction between strength of pre-NCLB accountability and stringency.

The coefficients are represented as Odds Ratios for logistic regressions in this study.

The strength effect represents the impact of moving from one level of pre-NCLB accountability strength to the next level (i.e.: Weak to Moderate).

The stringency effect represents the impact of a one percent increase in stringency.

The interaction effect is the differential effect of stringency on the various levels of pre-NCLB accountability. For example, as the strength of pre-NCLB accountability increases the negative impact of stringency on Elementary School teachers per 1000 significantly increased (-1.01 per level).

Figure 11: Student Choice and pre-NCLB Accountability Strength

Table 15: Comparing select outcomes between Public and Private School Teachers

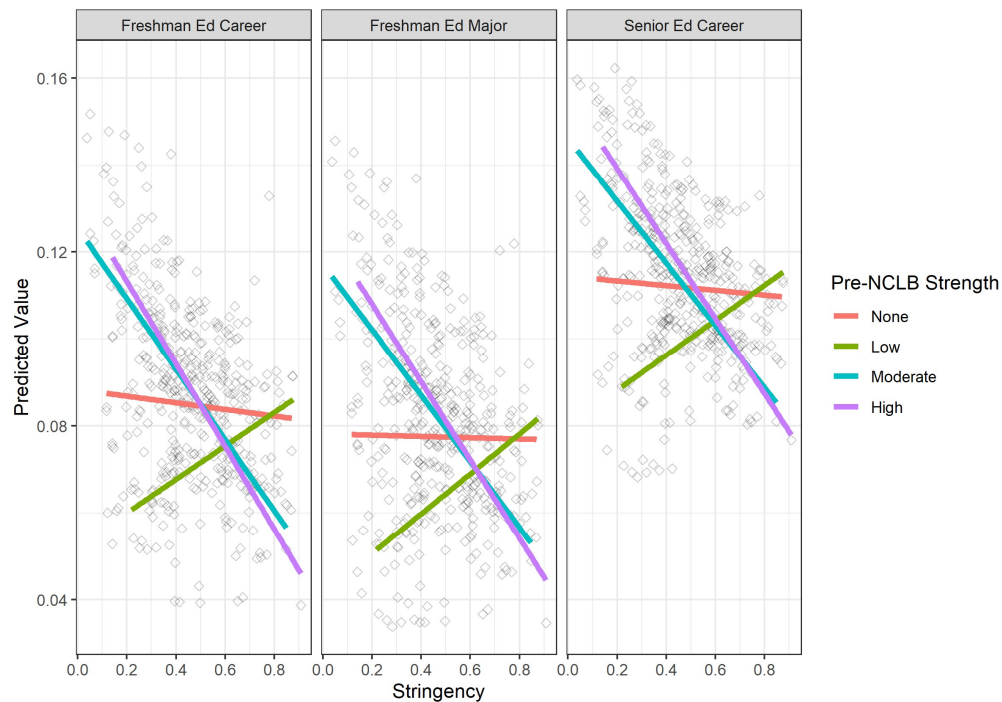|		Public	Private
Teacher Movement			
	Mover	1.366*	0.383
	Leaver	0.993	0.932
	Stayer	0.847	1.353
Teacher Perceptions			
	Control	0.023	0.007
	Salary Satisfaction	-0.116	0.013
	Admin Support	0.020	-0.026
	Colleague Support	0.019	-0.024

*p<.10

Teacher perception outcomes were estimated using a linear regression on simulated failure rates, state and year fixed effects, time-varying covariates, and a measure for the strength of recession.

Teacher Movement outcomes were estimated using a logistic regression with the same model specifications as above.

The coefficients for Teacher Movement are illustrated in Odds Ratios Teacher and the coefficients for Perceptions represent raw changes in teacher perception numbers.

The Public column represents the impact of a one percent change in stringency on public school teachers and the Private column represents the impact of a one percent change in stringency on private school teachers.

Table 16: Checking the Effect of lagging and leading treatment variables

		Leading		Lagged		Lag+3			
		String - 1	Stringency	String + 1	Stringency	String + 1	String + 2	String + 3	Stringency
Teacher Movement									
	Mover	2.067	0.662	4.393	0.313	21.542	0.200	2.052	0.156
	Leaver	1.301	0.763	1.194	0.833	0.888	1.350	0.780	1.061
	Stayer	0.595	1.425	0.393	2.151	0.204	1.941	0.772	2.770
Teacher Perceptions									
	Control	0.182	-0.159	-0.085	0.108	-0.165	-0.258	.336*	0.111
	Family Problems	0.562**	-0.544*	-0.169	0.186	0.022	-0.314	0.122	0.187
	Job Satisfaction	0.025	-0.034	-0.324	0.315	-0.183	-0.488	0.344	0.317
	Salary Satisfaction	0.318	-0.433	0.54	-.655*	0.8	-0.089	-0.169	-.657*
	Student Problems	.627***	-.649***	0.145	-0.168	0.184	-0.045	0.005	-0.168
	Admin Support	-0.097	0.117	-0.428	0.447	0.062	-.678*	0.187	0.448
	Colleague Support	-0.019	0.038	0.054	-0.034	0.434	-.736**	0.353	-0.032

*p<.10, **p<.05, ***p<.01

Teacher perception outcomes were estimated using a linear regression on simulated failure rates, leading and/or lagged simulated failure rates, state and year fixed effects, time-varying covariates, and a measure for the strength of recession.

Teacher Movement outcomes were estimated using a logistic regression with the same model specifications as above.

The coefficients for Teacher Movement are illustrated in Odds Ratios and the coefficients for Teacher Perceptions illustrate the raw change in a given outcome/perception score.

Stringency represents the impact of a one percent change in stringency on a given outcome.

String-1 illustrates the impact of a one percent change in stringency from the previous year

String+1 illustrates the impact of a one percent change in stringency from the following year

String+2 illustrates the impact of a one percent change in stringency from two years in the future

String+3 illustrates the impact of a one percent change in stringency from three years in the future

<div align="center">Appendix B: Equations</div>

One prior year leading treatment indicator

$$Y_{st} = \beta_0 + \beta_1 \ln(\overline{F}_{st-1}) + \beta_2 \ln(\overline{F}_{st}) + \mu_s + \alpha_t + XB_{st} + \varepsilon_{st}$$

Three prior years leading treatment indicator

$$Y_{st} = \beta_0 + \beta_1 \ln(\overline{F}_{st-1}) + \beta_2 \ln(\overline{F}_{st}) + \beta_3 \ln(\overline{F}_{st+1}) + \beta_4 \ln(\overline{F}_{st+2}) + \mu_s + \alpha_t + XB_{st} + \varepsilon_{st}$$

Lagged Model

$$Y_{st} = \beta_0 + \beta_1 \ln(\overline{F}_{st-1}) + \beta_2 \ln(\overline{F}_{st-2}) + \cdots + \beta_m \ln(\overline{F}_{st-m}) + \mu_s + \alpha_t + XB_{st} + \varepsilon_{st}$$

Appendix C: Determining State Accountability Stringency

Minimum Subgroup Size

The first rule taken into account when calculating whether or not a school was able to reach adequate yearly progress is the minimum number of students per subgroup by grade and subject to be held accountable for that subgroup. In general, this involves looking at whether or not a given subgroup has more students taking the assessment than a single number representing the minimum cutoff for that state and year, however, in some instances the minimum subgroup size cutoff was a percentage of total number of students within that grade. The sizes of the minimum subgroup cutoff range from the lowest minimum being Maryland with an accountable subgroup size of 5 to the highest minimum being California with an accountable subgroup size of 100, with a majority of the other states falling in between with a minimum cutoff ranging from 30 to 40 students. If a subgroup has more students taking the assessment than the minimum cutoff for that grade and subject, than they will count as an accountable subgroup, while if there were less students taking the assessment for that grade and subject than the subgroup will not count as an accountable subgroup.

Proficiency Thresholds/Annual Measurable Objectives for each Subject/Grade

Much of the variation across in accountability rules across the states and the years is in the setting of the proficiency thresholds or annual measurable objectives (AMO) for each subject and grade. The annual measurable objectives represent the minimum level of proficiency or percent proficiency that a subgroup needs to reach in order to reach adequate yearly progress. The overall goal of No Child Left Behind was to reach 100 percent proficiency in all states in both reading and math by the year 2014, but the states were able to set their proficiency thresholds before 2014 in any way they deem necessary, oftentimes with small yearly increments

and other times with the proficiency threshold staying steady for 2 or 3 years and then making a larger jump. Many states also started their proficiency thresholds at a different level, with some states beginning at a lower level having to face a more drastic increase in proficiency to reach the 100 percent goal and some states starting at a higher level and having to maintain that high level to reach the end goal. The table below illustrates the minimum, maximum and average proficiency threshold with standard deviation for 2002-2003, 2006-2007, and 2010-2011 for each subject for grades 4 and 8.

Table C1

Proficiency Thresholds for Math and ELA

		Math		English/Language Arts	
		Grade 4	Grade 8	Grade 4	Grade 8
2003	Min	9.3	7	13.6	13.6
	Max	75.9	74.6	76.9	77
	Mean (SD)	44.1 (19.2)	37.8 (18.3)	52.8 (17.7)	48.1 (16.7)
2007	Min	26.5	20	24.4	24.4
	Max	83.6	79	83	83
	Mean (SD)	56.9 (14.7)	52.6 (15.2)	63.3 (14.5)	61.2 (14.8)
2011	Min	40	40	49	49
	Max	94.5	92.6	94.2	93.6
	Mean (SD)	75.1 (10.8)	73.7 (10.5)	78.2 (10.3)	77.7 (10.2)

This table illustrates the great variation in thresholds across the years, however, this variation did prove to become less drastic as the implementation of No Child Left Behind grew mature.

Additional Academic Indicators

The next main rule that clarifies whether or not a school makes adequate yearly progress is the additional or other academic indicator (OAI), which is one extra indicator chosen by the state and that receives federal approval. At the high school level this additional indicator is the

graduation rate and whether or not the high school reaches the required graduation rate threshold, however, the additional indicator at the elementary and middle school levels is much more varied. While states varied in their chosen additional indicator, a majority (72.5%) of the states used attendance rates as the additional academic indicator with required attendance rate cutoffs ranging from 76% to 95% with 65% of the states using attendance rates allowing an improvement in attendance rates to act as a means to passing the additional indicator if the required rate cutoff is not met. Other states used scores or a combination of scores on additional subjects (Science, Writing, History) as an additional indicator, while other states set targets for the improvement of the percentage of students at either the very bottom level of proficiency (Below Basic) or at the highest level of proficiency (Advanced), and finally, other states used retention rates or unexcused absences as the additional indicator. Each school is

To calculate the stringency measure and to check if a school made adequate yearly progress, each school measures against the academic indicator threshold or checks for improvement over the previous year, if the improvement rule is a part of the accountability plan for the state.

Use and Size of Confidence Intervals

One of the more common exemption rules in use across the states was the use of a confidence interval, usually within the 95% to 99% range, around the proficiency threshold or annual measurable objective target. The use of a confidence interval allows a state to adjust the target downwards depending upon the size of the confidence interval, the initial annual measurable objective, and the number of students within the subgroup. For example, without the use of a confidence interval, a subgroup with 50 students facing an annual measurable objective of 55% that only has 52% of students that are proficient would fail to reach adequate yearly progress. However, applying a 95% confidence interval around the target would essentially

lower the target below 52% allowing the previously failing subgroup to pass the new threshold and meet the performance requirement for adequate yearly progress.

Use of Safe Harbor

The next exemption rule is the Safe Harbor rule, utilized by nearly all states, that essentially allows a subgroup to reach their adequate yearly progress performance requirement if there was a decrease in the percentage of students not meeting the proficiency threshold from the previous year and also progress made on meeting the additional academic indicator. Although there are slight variations in the implementation of the Safe Harbor rule mentioned above, it generally calls for a 10% decrease in the percentage of students not meeting the proficiency threshold from the previous year with a small number of states allowing for more previous years' performance in the calculation. Fundamentally, this allows a subgroup to make progress or improvement towards the proficiency goal, even if they are not fully able to reach the established annual measurable objective in that given year.

Use of a Confidence Interval around Safe Harbor

In addition to the Safe Harbor rule, around one third of the states utilized a 75% confidence interval around the Safe Harbor target calculations when calculating whether or not a subgroup made their performance requirement. The use of the 75% confidence interval is similar to the earlier use of the confidence interval and provides an even greater chance for a subgroup to make their target without reaching the original Annual Measurable Objective cutoff.

Use of Multiple Years to Average Performance

Much like Safe Harbor which looked at past performance, the multiple year averaging rule uses multiple years of student performance to check if a subgroup reaches the required performance target. This rule, utilized by around one third of the states, allows for an average of

multiple years to calculate student performance and also the use of a confidence interval in combination with the average of student performance. The averaging calculation involves adding the current year's performance to the previous 1 or 2 years of performance, depending on whether the rule in the state and year calls for the use of 2 or 3 years, and dividing by the total number of years. The use of the multi-year averaged performance allows a subgroup to reach the performance requirement if the average performance is greater than the original proficiency threshold or greater than the confidence interval adjusted target.

Use of Performance Indices and Proficiency Indices

As was discussed earlier in the literature review, a small number of states utilized performance (13) and proficiency (6) indices in their determination of AYP, a strategy that essentially boosts aggregate school performance numbers. The use of a performance index places weights across various performance levels and does not simply count the percentage above proficient, fundamentally providing partial credit to students who did not reach the proficiency threshold. Proficiency indices, on the other hand, provide a way to account for disparate proficiency thresholds across grades or grade spans by weighting the performance employing enrollment numbers and checking for proficiency across the school and not just within one grade.

The proficiency index rule was not included in the stringency calculation, because of the need for performance scores for multiple consecutive grades, which is not possible when utilizing NAEP assessments. However, the performance indices is included in the stringency calculation through the use of a mapping technique similar to the test difficulty and test equivalency calculation, which is a part of a more detailed discussion in a subsequent section. Essentially, a comparison occurs between the distributions of NAEP scores and statewide

accountability assessments scores within the state with the performance index, creating a NAEP

cutoff for each performance level. A cutoff is found in the statewide accountability assessment

for the given performance level, for instance in Alabama if 78% of students scored above the

partially meeting standards or Level II performance level on their standardized assessment, than

the cutoff for this performance level within the NAEP sample would be set at the same percentile

in the NAEP distribution. Each school than calculates their performance index score based on the

state and year performance index rules, which was then scaled to fit with the proficiency

threshold set by the state and year and is then compared against that threshold to check the

adequate yearly progress performance requirement.[1]

Table C2

Overview of the Performance Index Rule

State	Performance Index Rule
Alabama	Weights students at or above proficient 1.0; partially meets standards at .5; and those not meeting standards at zero.
Idaho	Allocates 100 points for performance at or above proficient and 50 for basic. A school's index score is the average of all student index points assigned to the school.
Massachusetts	Allocates 0 points for warning/failing low; 25 for warning/failing high; 50 for needs improvement low; 75 for needs improvement high; and 100 for proficient or advanced
Minnesota	Allocates One half point for each student in Level 2 (partially proficient) and One full point for each student in Level 3 (meeting or exceeding proficiency).
Mississippi	Weights the proportion of students scoring in the proficient or advanced by 1.0, the proportion scoring in basic by 0.5, and the proportion scoring in minimal by zero.
New Hampshire	Allocates 100 points for performance at or above proficient, 80 for the upper portion of partially proficient, 60 points for the lower, 40 points for the upper portion of substantially below proficient, 20 points for the lower, and zero for no score or response that fell with a "guessing" parameter.

---

[1] The Appendix contains a table of the performance indices rules for each of the states that utilized a performance index.

New York	The NY performance index involves computing the percentage of students scoring at Levels 2 (basic proficiency), 3 (proficient), and 4 (advanced), and twice the percentage of students scoring at Levels 3 and 4 only, divided by all continuously enrolled students.
Oklahoma	Allocated 1 point for Unsatisfactory, 2 points for Limited Knowledge and 3 points for Proficient and Above. Converted to a scale score between 20 and 80.
Pennsylvania	Allocated 1 full point for Proficient and above, .8 for High Basic, .6 for Low Basic, .4 for High Below Basic, .2 for Low Below Basic and 0 for not tested.
Rhode Island	Allocates 100 points for Proficient and above, 75 for Below Proficient, 50 for Substantially Below Proficient and Upper Half of Scale Range, 25 for Substantially Below Proficient and Lower Half of Scale Range, and 0 for not tested.
South Carolina	Allocates 100 points for Proficient or Advanced; 75 for Basic; 50 for Below Basic 2; and 25 for Below Basic 1
Vermont	Performance in the lower half of this lowest level has a value of 125 points; 250 in the upper half; 375 in the just below proficient level; and 500 at the proficient and advanced levels.
Wisconsin	Awards one point for all proficient and advanced scores, one-half for scores at basic, and none for below basic.

Table C3

Three State Example

	Mississippi	North Carolina	Massachusetts
Minimum Subgroup Size	40	40	40
Annual Measurable Objective – Math Grade 4	70	89	84
Annual Measurable Objective – Math Grade 8	66	89	84
Annual Measurable Objective – ELA Grade 4	67	72	90
Annual Measurable Objective – ELA Grade 8	66	72	90
Performance Index	Yes	No	Yes
Safe Harbor (with Confidence Interval?)	Yes (No)	Yes (No)	Yes (No)
Confidence Interval (Size)	Yes (99%)	Yes (95%)	Yes (95%)
Standardized Test Difficulty Ranking: Grade 4 Math, Grade 8 Math, Grade 4 ELA, Grade 8 ELA	22, 34, 7, 9	27, 42, 19, 23	1, 1, 1, 18
Other Academic Indicator (Grades 3 – 8)	Attendance (90%)	Attendance (90%)	Attendance (92%)
Stringency (School Failure Rate)	30%	54%	88%

Appendix D

School and Staffing Survey Overview

Scale	Questions	Scale Score (SD)	Average Scale Score (SD)
Classroom Control (alpha = .76) This scale measures the amount of control a teacher reported having to make autonomous decisions in his or her classroom and asked teachers about:	Teacher has control in selecting textbooks and other instructional materials	2.9 (.97)	3.4 (.51)
	selecting content, topics, and skills to be taught	3.0 (.95)	
	Teacher has control in selecting teaching techniques	3.6 (.61)	
	Teacher has control in evaluating and grading students	3.6 (.59)	
	Teacher has control in disciplining students	3.4 (.70)	
	Teacher has control in determining the amount of homework to be assigned	3.65 (.62)	
Student Problems (alpha = .79) This scale measures the amount of student behavior problems reported by the teacher	Student tardiness is an issue	2.2 (.85)	2.0 (.63)
	Student absenteeism is an issue	2.6 (.86)	
	Students cutting class is an issue	1.5 (.74)	
	Students dropping out is an issue	1.5 (.76)	
	Student apathy is an issue	2.4 (1.00)	
Family Problems (alpha = .84) This scale measures perceived problems associated with poverty and lack of family support and asked teachers about:	Lack of parental involvement is an issue	2.7 (.96)	2.5 (.75)
	Students coming to school unprepared is an issue	2.8 (.91)	
	Poor student health is an issue	2.0 (.80)	
	Poverty is an issue	2.6 (.95)	
Job Satisfaction (alpha = .62) The job satisfaction scale captures teacher reports of satisfaction with their particular school and teaching assignment	I am satisfied with being a teacher at the school	3.3 (.85)	3.0 (.60)
	Teachers at this school are a satisfied group	3.1 (.80)	
	I like the way things are run at the school	3.0 (.84)	
	I am satisfied with my salary	2.2 (1.00)	

Administrative Support (alpha = .83) The administrative support scale captures the level of satisfaction teachers have with the support received from the administration	Administrator behavior being supportive and encouraging	3.2 (.88)	3.2 (.71)
	Administrator's ability to enforce school rules and back up the teacher when needed	3.3 (.86)	
	Ability to communicate clearly the mission of the school	3.3 (.84)	
	Recognize members for doing good work	2.9 (.91)	
Colleague Support (alpha = .74) The colleague support scale captures the level of satisfaction teachers have with the support received from their colleagues	Teachers consistently enforce rules, even for students not in their classrooms	2.9 (.90)	3.1 (.67)
	Sharing a common belief about the mission of the school	3.2 (.73)	
	Coop with other teachers	3.2 (.82)	
Career Satisfaction (alpha = .57) The career satisfaction scale captures teacher reports of satisfaction with the career of teaching	How long do they plan on staying in teaching	3.3 (.79)	3.6 (.85)
	Would they choose teaching if they had to do it over	3.8 (1.2)	
Burnout (alpha = .77) This scale measures teacher burnout, asking teachers to respond to statements about their enjoyment of teaching and their desire to leave their school or the profession.	The stress and disappointments involved in teaching at this school aren't really worth it	3.2 (.81)	3.1 (.67)
	The stress and disappointments involved in teaching at this school aren't really	3 (.94)	
	I think about transferring to another school.	3.1 (.98)	
	I don't seem to have as much enthusiasm now as I did when I began teaching.	2.9 (1.00)	
	I think about staying home from school because I'm just too tired to go.	3.3 (.88)	
Burnout/Career Satisfaction (alpha = .80)			3.3 (.64)
Burnout/Career/Job Satisfaction (alpha = .83)			3.2 (.56)

Appendix F: Pre and Post NCLB Variables

Table 17: Distribution of States by Pre-NCLB Accountability Strength

None	Low	Moderate	High
Arizona	Alaska	Connecticut	Alabama
Hawaii	Arkansas	Georgia	California
Idaho	Delaware	Kansas	Colorado
Iowa	District of Columbia	Michigan	Florida
Maine	Massachusetts	Nevada	Illinois
Minnesota		Oklahoma	Indiana
Mississippi		Oregon	Kentucky
Missouri		Rhode Island	Louisiana
Montana		South Carolina	Maryland
Nebraska		Tennessee	New Mexico
New Hampshire		Vermont	New York
New Jersey		Virginia	North Carolina
North Dakota		Wisconsin	Texas
Ohio			West Virginia
Pennsylvania			
South Dakota			
Utah			
Washington			
Wyoming			
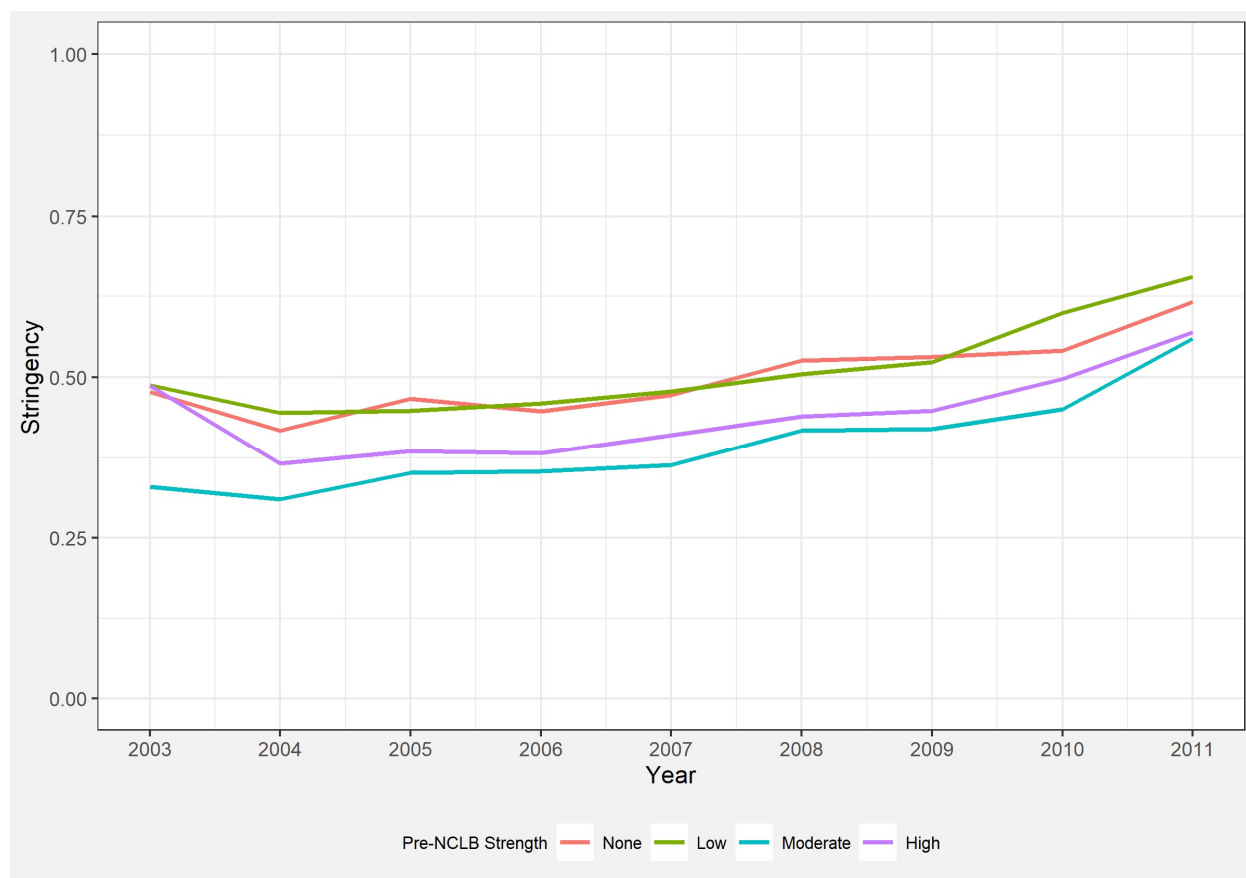
Figure 12: Stringency and Pre-NCLB Accountability Strength

Table 18:Coding Rules for Previous Accountability Strength

State	Previous Accountability Strength	Type of Policy (Hanushek and Raymond)	School Repercussions (Carnoy and Loeb)	Accountability Type (Lee and Wong)
Alabama	High	Consequential	Strong	Strong
Alaska	Low		Weak	Weak
Arizona	None			
Arkansas	Low	Consequential	None	Weak
California	High	Consequential	Strong	Moderate
Colorado	None			
Connecticut	Moderate	Consequential	Weak	Moderate
Delaware	Low	Consequential	None	Weak
District of Columbia	None			
Florida	High	Consequential	Strong	Strong
Georgia	Moderate	Consequential	None	Moderate
Hawaii	None			
Idaho	None			
Illinois	High		Moderate	Strong
Indiana	High	Report_Card	Moderate	Strong
Iowa	None			
Kansas	Moderate	Report_Card	Weak	Moderate
Kentucky	High	Consequential	Strong	Strong
Louisiana	High	Consequential	Moderate	Strong
Maine	None			
Maryland	High	Consequential	Strong	Strong
Massachusetts	Low	Consequential	Implicit only	Weak
Michigan	Moderate	Consequential	Weak	Moderate
Minnesota	None			
Mississippi	None			
Missouri	None			
Montana	None			
Nebraska	None			
Nevada	Moderate	Consequential	Weak	Moderate
New Hampshire	None			
New Jersey	None			
New Mexico	High	Consequential	Moderate to strong	Strong
New York	High	Consequential	Strong	Strong
North Carolina	High	Consequential	Strong	Strong
North Dakota	None			
Ohio	None			
Oklahoma	Moderate	Consequential	Weak	Moderate
Oregon	Moderate	Consequential	Weak to Moderate	Moderate
Pennsylvania	None			
Rhode Island	Moderate	Consequential	Weak	Moderate
South Carolina	Moderate	Consequential	Moderate	Moderate
South Dakota	None			
Tennessee	Moderate	Consequential	Weak	Moderate
Texas	High	Consequential	Strong	Strong
Utah	None			
Vermont	Moderate	Consequential	Weak	Moderate
Virginia	Moderate	Consequential	Weak to Moderate	Moderate
Washington	None			
West Virginia	High	Consequential	Strong	Moderate
Wisconsin	Moderate	Consequential	Weak to Moderate	Moderate
Wyoming	None			

	Pre-NCLB		Post-NCLB		
	Mean	Standard Deviation	Mean	Standard Deviation	P
State Population Characteristics					
State Population	5701744.63	6426006.79	5919267.35	6628288.53	0.825
Poverty Rate	11.75	3.22	12.73	3.28	0.043
Unemployment Rate	5.65	1.05	6.25	2.32	0.067
% with Bachelor's Degree	26.2	4.61	26.84	4.42	0.38
% Republican (Governor's Affiliation)	55	50	49	50	0.39
Teacher Characteristics					
% Less than 35 Years Old	16.03	2.55	15.11	3.35	0.063
% White	84.13	16.03	84.95	16.2	0.743
% Black	7.69	12.29	7.07	11.97	0.736
% Hispanic	3.15	5.15	3.63	5.65	0.575
Pupil Teacher Ratio	15.5	2.32	15.14	2.65	0.352
Student Characteristics					
% Black Students	15.72	16.48	15.71	15.91	0.997
% White Students	67.71	19.62	64.24	19.64	0.232
% Hispanic Students	9.89	11.62	12.25	12.62	0.202
% of Students on FRPL	29	11	33	10	0.023