School Accountability and Principal Mobility: How No Child Left Behind Affects the Allocation of School Leaders [*]

Danielle Li[†]

This draft: October 16, 2012

Abstract

The move toward increased school accountability may substantially affect the career risks that school leaders face without providing commensurate changes in pay. Since effective school leaders likely have significant scope in choosing where to work, these uncompensated risks may undermine the efficacy of accountability reforms by limiting the ability of low-performing schools to attract and retain effective leaders. This paper empirically evaluates the economic importance of principal mobility in response to accountability by analyzing how the implementation of No Child Left Behind (NCLB) in North Carolina affected principal mobility across North Carolina schools and how it reshaped the distribution of high-performing principals across low- and high-performing schools. Using value-added measures of principal performance and variation in pre-period student demographics to identify schools that are likely to miss performance targets, I show that NCLB decreases average principal quality at schools serving disadvantaged students by inducing more able principals to move to schools less likely to face NCLB sanctions. These results are consistent with a model of principal-school matching in which school districts are unable to compensate principals for the increased likelihood of sanctions at schools with historically low-performing students.

[*]I am very grateful to Daron Acemoglu, Leila Agha, Josh Angrist, David Autor, Esther Duflo, David Figlio, Brigham Frandsen, Michael Greenstone, Matt Notowidigdo, Amanda Pallais, Parag Pathak, and numerous participants at the MIT Labor Lunch for helpful comments and suggestions. All errors are my own.

[†]Northwestern University, Kellogg School of Management

1 Introduction

The No Child Left Behind Act of 2001 (NCLB) has motivated a vast research program studying the effects of test-based accountability on student performance in U.S. public schools. Though the degree to which test score gains documented at schools under the threat of sanction reflect durable improvements in learning versus strategic behaviors is the subject of debate, one finding of this literature is unambiguous: test-based accountability has significantly changed the incentives and working conditions of teachers and principals.[1] For instance, Reback, Rockoff, and Schwartz (2011) find that NCLB accountability pressures lead untenured teachers to work longer hours and feel less secure in their jobs. Yet despite increased scrutiny at disadvantaged schools, principal pay did not always adjust to compensate. This relative change in the risk-reward structure of low- versus high-performing schools raises the concern that NCLB might induce effective principals at low-performing schools—who may have the option of working elsewhere—to differentially depart these schools.

This paper provides the first quantitative evidence that I am aware of on this possibility. A recent study by Jackson shows that high-poverty schools facing competition from charter school entry are able to retain their best teachers by increasing salaries, but still have a harder time filling new vacancies (Jackson, 2012). In the absence of wage flexibility, however, the introduction of accountability policies may have even more detrimental effects. My results indicate that in evaluating NCLB's impact on students, it is important not only to consider short term test score gains but also the long-term allocative effects of increasing accountability without increasing compensation.

The labor market choices of educators is a critical channel by which NCLB may affect school quality in the long run. An influential body of work demonstrates that teacher and principal

[1] Drawing on data from both NCLB and smaller state and district-based accountability programs, studies of the effect of accountability on test scores broadly conclude that accountability programs can raise test scores at poorly-performing schools. Figlio and Rouse (2006), West and Peterson (2006), Rouse et. al. (2007), Chiang (2008), Krieg (2008), Neal and Schanzenbach (2007), and Dee and Jacob (2009) all find test score gains of some kind. The nature of these gains is the subject of more debate. Rouse et. al. (2007) and Chiang (2008) find persistent gains in math test scores under Florida's state accountability system, but West and Peterson (2006) show that these test score gains do not carry over to NCLB. Krieg (2008), and Neal and Schanzenbach (2007) do find that NCLB increases test scores, but raise concerns that gains are concentrated in the middle of the ability distribution, suggesting that schools ignore low- and high-achieving students in favor of marginal students. Figlio and Getzler (2002), Jacob (2005), and Reback (2006) show that schools remove poorly-performing students from testing pools by reclassifying low-achieving students into special education and Figlio (2006) documents a similar phenomenon where poorly-performing students are subjected to longer disciplinary suspensions near testing dates. Jacob and Levitt (2003) study teacher cheating in pressured schools, and Figlio and Winicki (2005) document calorie inflation.

quality is a major determinant of student learning and that assigning a student to a good educator can matter more for learning than reducing classroom size or increasing classroom resources.[2] Principals, in particular, play an important role in recruiting, developing, and retaining good teachers.[3] Yet unlike the number of computers or teachers per student, the quality of a school's staff cannot simply be assigned. Rather, teachers and principals make choices both about where to work and how much effort to exert.

Loeb, Kalogrides, and Horng (2010) and Cullen and Mazzeo (2007) show that principals have preferences over the types of schools they serve are motivated by the opportunity to change schools. In Cullen and Mazzeo's model, career concerns can improve academic performance by creating a competitive environment in which principals exert effort even absent explicit performance bonuses or sanctions. Yet, the effect of increased accountability in such a competitive labor market can create unintended consequences for equity. If teachers and principals change jobs in response to the labor market incentives created by NCLB, then NCLB's initial effect on test scores may not reflect its full effects in equilibrium. In particular, labor market sorting may erode the efficacy of NCLB in achieving its stated objectives, which affirm the value of improving access to high-quality schools for disadvantaged children.

By requiring schools to meet the same proficiency targets regardless of prior student performance, NCLB creates wide variation in the likelihood that a school misses performance targets based on factors, such as student demographics, beyond a principal's control.[4] At the same time,

[2]Aaronson, Barrow, and Sander (2003), Rockoff (2004), and Hanushek, Kain, and Rivkin (2005), for instance, study teacher value-added. Branch, Hanushek, and Rivkin (2012) estimate principal value-added and find significant variation in quality, especially among principals at high-poverty schools. Early studies of principals include Eberts and Stone (1988) and Ballou and Podgursky (1995) who study predictors of principal effectiveness.

[3]Knapp et al, Plecki et. al., Portin et. al., and Copland and Boatright (2006, Wallace Foundation Report) argue that effective principals are able to develop leadership potential among teachers. Jacob and Lefgren (2005) highlight principals' roles in assessing teacher quality, and Jacob (2010) examines the role of principals in firing teachers. Rockoff et. al. (2011) provide more evidence that principals play and important role in evaluating and improving teacher performance and Branch, Hanushek, and Rivkin (2012) provide indirect evidence that principal value-added is correlated with higher turnover among poorly performing teachers. A recent literature on the determinants of school performance find that that school-wide factors that principals maintain—personnel policies, frequent teacher feedback, a commitment to No Excuses policies, for example—are important for student learning (Loeb, Kalogrides, and Beteille, 2011; Dobbie and Fryer, 2011; Angrist, Pathak, and Walters, 2012).

[4]In the first year of NCLB in North Carolina, where my analysis takes place, the passing thresholds for reading and math were set at, respectively, 68.9 and 74.6 percent proficiency, well above levels typical at schools serving low-income and minority children. Thus, a principal of a school with poorly-performing students would almost surely fail to meet these performance targets, known as Adequate Yearly Progress (AYP), in the first year, and would be subjected to increased administrative burdens as well as to an increased likelihood of facing sanctions in later years. Conversely, a principal of a school with high-performing students is likely to pass AYP almost regardless of his or her actions.

principal salaries, which continue to depend almost entirely on education and experience, have not differentially adjusted to compensate.[5] Though school districts do provide supplements above the standard salary scale, I show that supplements in my sample did not comparatively increase for principals at poorly-performing schools. Thus, NCLB represents a significant, and largely uncompensated change in the risk and amenities associated with working at disadvantaged schools.

Existing studies of the effect of accountability on teacher labor markets have reached mixed conclusions: Clotfelter et. al. (2004) find that accountability increases turnover at poorly-performing schools, but Boyd et. al. (2008) find that turnover decreases among teachers whose students are subject to testing. This literature does not address, however, whether or in which direction turnover matters for student performance.[6] Increased turnover might signal either that accountability makes it harder to retain effective teachers or that accountability makes it easier to dismiss ineffective ones. Turnover may not affect school quality at all if teachers or principals do not matter for student performance or if it does not change the ultimate composition of school staff.

The key contribution of this paper is to use outcome-based measures of principal quality to examine whether NCLB accountability leads principals to seek less demanding jobs. By tying mobility to quality, I can answer a rich set of questions regarding the impact of NCLB on principals: which types of principals are more likely to switch schools, what kinds of schools do they move to, and what happens to the distribution of principal quality across schools?

To perform this analysis, I estimate principal quality in the period prior to the implementation of NCLB by extracting principal value-added from student test scores. Next, I use variation in school demographics prior to the adoption of NCLB to measure the likelihood that a school will be subject to sanctions. I examine the impact of NCLB on the distribution of principal effectiveness, as well as on changes in mobility patterns, under the assumption that NCLB's accountability provisions should be more binding for principals of schools that are more likely to miss performance

[5] A small literature looks at principal pay and incentives. Billger (2007) uses cross-sectional comparisons to show that district sanctions against a school are associated with lower principal pay and mixed results for graduation and retention. Lavy (2008) uses difference-in-differences to estimate the impact of an Israeli program increasing principal pay by 50% and finds significant, though small, gains in student test scores and subjects taken. Besley and Machin (2009) use UK data to show that principal pay and retention responds to performance: pay is linked to publicly observable performance measures and poorly-performing principals face a higher chance of replacement.

[6] A recent paper by Hanushek and Rivkin tie teacher value-added to turnover and find that teachers who leave urban schools tend to be worse than the ones who stay. Their paper focuses primarily on the mobility choices of teachers early in their career as they discover their aptitude for teaching. I instead focus on the mobility decisions of seasoned educators in response to accountability.

targets based on their pre-period demographics.

Consistent with the existing literature, I find that principals matter for performance and that their effectiveness varies significantly across schools. I show that after NCLB high-ability principals at schools more likely to face sanctions for missing performance targets, known as Adequate Yearly Progress (AYP), disproportionately move to schools less likely to face sanctions. These changes in the assignment of schools to principals translate into economically substantive declines in principal effectiveness at schools serving disadvantaged student populations. As a result of NCLB, a standard deviation increase in the likelihood that a school fails AYP leads to a one-third standard deviation decrease in average principal effectiveness, as measured by value-added to students' math test scores. These findings are consistent with a model of principal-school matching in which asymmetric changes in the probability that a principal will face performance sanctions, when not fully compensated by changes in pay, lead principals to prefer schools where they are less likely to face sanctions.

A potential limitation of this approach is that principal value-added is estimated only for principals who switch schools between 1995 and 2002, leading to a selected and relatively small sample of principals. Specification checks, as discussed in the empirical appendix, however, indicate that this selection is unlikely to bias estimates of the *differential* impact of NCLB on high- and low-performing schools.

In the next section, I discuss the implementation of NCLB in North Carolina. Section 3 outlines a model of principal-school matching under accountability. Section 4 outlines my econometric methods and estimating equations. Section 5 describes the data and sample construction. Section 6 presents results and Section 7 concludes.

2 Institutional Context

No Child Left Behind was signed into federal law in January 2002 with the goal of enabling all children access to a high-quality education as measured by universal proficiency in math and reading by 2014. It mandated annual testing in these subjects for all students in grades 3 through 8, and at least once during high school, starting in the 2002-2003 school year. Under NCLB, schools are designated passing or failing depending on whether they make a performance target known as Adequate Yearly Progress (AYP). Schools are divided into 9 demographic subgroups and AYP requires that students in each subgroup with over 40 members reach a particular threshold for reading and math scores.[7] If only one subgroup fails to make this target, the entire school is declared failing. Starting in 2003-04, NCLB was amended to include a "Safe Harbor" provision, wherein schools could also make AYP by demonstrating a 10% improvement in scores for *every* subgroup that still falls below the performance target. In practice, however, very few schools passed AYP under Safe Harbor because it was difficult to improve test scores in among every low-performing demographic subgroup.

Sanctions associated with failure to make AYP varies across schools and, in particular, depends on whether 1) a school receives federal Title I funds and, if not, whether the school is located in a district which receives Title I funds. Regardless of funding status, NCLB requires that report cards comparing the performance of all schools be made public. Additionally, after two consecutive years of failing to make AYP in the same subject, all schools are required to develop a School Improvement Plan describing strategies that the school will use to meet future performance targets.

The primary bite of NCLB, however, comes at schools which receive federal Title I funds (approximately 50% of schools). Schools are eligible for Title I funding if they serve a large number or high percentage of poor students. For these schools, NCLB created a schedule of sanctions based on the number of consecutive years a school fails to meet AYP in the same subject. AYP designations are determined in the spring and sanctions apply for the following school year:[8]

- First year: There are no official sanctions for the next year, but parents are notified that their child's school is failing.

- Two consecutive years: The school enters the first year of "Title I Improvement" the following school year. In this phase, schools must enable parents to send their children to a non-failing school in the district, unless the school is in a pilot district offering supplemental educational services as the first year option.

- Three consecutive years: The school enters Year 2 of Title I Improvement at the beginning of

[7] The subgroups are 1) White; 2) Black; 3) Hispanic; 4) Native American; 5) Asian/Pacific Islander; 7) Multiracial; 7) Free/Reduced Price Lunch Students; 8) Limited English Proficient Students; and 9) Students with Disabilities.

[8] For more details, see the North Carolina Public Schools' NCLB overview: http://www.ncpublicschools.org/nclb/abcayp/overview/ayp

the next school year and continues to implement school choice and supplemental educational services.

- Four consecutive years: The school enters Year 3 of Title I Improvement at the beginning of the next school year. School choice and supplemental educational services continue. In addition, the district can pursue "corrective action," meaning that it can replace the principal and teachers, and restructure the curriculum.

- Five consecutive years: The school enters Year 4 of Title I Improvement at the beginning of the next school year. In addition to the sanctions above, the school must devise a contingency plan for restructuring, where the school can be closed, reopened as a charter, privatized, or be taken over by the state.

- Six consecutive years: The school enters Year 5 of Title I Improvement at the beginning of the next school year. Restructuring plans can be implemented.

Non-Title I schools that are located in districts which receive Title I funds (almost all non-Title I schools in my sample) can be sanctioned at the district level if their district fails to meet AYP as a whole. In these cases, however, the primary accountability falls on the superintendent.

Prior to the enactment of NCLB, there were no federally-mandated standards that governed accountability and testing in US public schools. States (and to a lesser extent, districts) had significant purview in designing their own standards for school performance monitoring and accountability. In practice, however, while most states conducted annual testing, very few had explicit consequences associated with poor performance.

North Carolina, the setting for this analysis, was a notable exception because it already had an accountability program in place prior to the introduction of NCLB. Though that program, the ABCs of Growth, was an early model for NCLB, its benchmarks and requirements differed substantially. In particular, the ABCs emphasized performance as measured by gains in student performance. The ABCs were first implemented for K-8 students in the 1996-97 school year and, initially, schools were given one of the following four designations: 1) Exemplary, for schools whose average test score gains exceeded expected gains by over 10%; 2) Meets Expectations, for schools whose gains meet expectations but which do not exceed them by 10%; 3) No Recognition, for schools that do not meet growth standards, but whose students are more than 50% proficient; 4) Low Performing, for schools that do not meet growth standards, and whose students are less than 50% proficient.[9] Teachers at schools in the top two categories received small bonuses ($1,500 and $750, respectively).

When considering how the implementation of NCLB may have affected principals, the relevant benchmark is how NCLB changed the perception of sanctions relative to the ABCs, not relative to no accountability at all. Importantly, there were no explicit sanctions associated with poor performance and, in practice, the ABC designations were relatively non-binding: in most years, fewer than 1% of schools were designated Low-Performing. In fact, in the 2001-02 school year, on the eve of NCLB's implementation, only 7 schools, or 0.34% of all schools, failed to meet ABC standards. In contrast, in 2002-03, 53% of schools failed to make AYP in the first year, and in 2004-05, almost 10% of schools were subject to official NCLB sanctions. Moreover, school performance on the ABCs was only weakly correlated with AYP performance; 44% of ABC Exemplary schools failed to make AYP and 73% of schools meeting growth expectations under the ABCs failed to make AYP. In particular, because ABC designations are based on gains in scores, student demographics are a much stronger predictor of AYP performance; the percentage of white and free lunch students, for instance, explains over 28 percent of variation in AYP status, as opposed to just over 6 percent of variation in ABC status. Because schools with many disadvantaged students are significantly more likely to pass ABCs, the implementation of NCLB particularly affects principals of these schools, relative to the ABCs. In Section 6.4, I explore whether principals at schools that were "shocked"—those that met ABC's growth standards but not NCLB's level standards—were particularly affected. I show, in particular, that principal quality differentially fell at schools which met ABC expectations in 2002, but which failed a large percentage of AYP targets in the first year of NCLB.

Institutional features of the market for school principals play a significant role in shaping how principals respond to these changes in accountability pressure. Principal salaries are set by a statewide schedule that is primarily a function of principal experience, education, and school size. Principals receive the same state wage regardless of school quality, conditional on size. Further, regardless of ability, principals with the same education and experience also receive the same wage,

[9] Expected gains were calculated by regressing student level gains on student characteristics using 1994 data and then applying the estimated coefficients to data from future years. For more details, see Ladd and Walsh (2002).

even though studies have found no relationship between principal education and ability, and little relationship between experience and ability beyond the first two years. School districts may provide additional salary supplements for principals, usually around 10% of total pay, which does vary from district to district, but I do not find evidence that districts systematically compensate principals for the quality of the schools at which they work (Appendix Table A).

In North Carolina, principals work on four-year contracts and are not unionized. This increases accountability pressures in the early years of the Title I Improvement phase because districts do not need to wait until the restructuring phase, when principal replacement is explicitly endorsed under NCLB, to act on information about school performance revealed through testing. Further, in contrast to unionized states, there are no strict seniority preferences in hiring; this means that if principals do respond to accountability pressures, mobility may be more related to performance measures as opposed to measures of tenure or experience.

When a principal vacancy is created, districts post this open position and solicit applications.[10] From the pool of applicants, schools pick finalists who are then invited for onsite interviews with the district, school, and school board. Even though principals are officially employed by their local school district, individual schools make offers to candidates and candidates may receive offers from multiple schools in the same district. Importantly, superintendents cannot explicitly transfer principals to other schools within the district and this limits the extent to which principal moves do not reflect optimization by principals given their choice set.

3 Theoretical Framework

NCLB imposes sanctions on schools and their leadership when students fail to achieve a certain level of proficiency on annual tests. Because school demographics strongly predict test scores, NCLB changes the implicit costs of working with students from disadvantaged backgrounds. The following simple model illustrates how principal preferences translate into principal-school allocations and looks at the effect of accountability on these allocations. I ignore the effect of accountability on principal retirement or firing in order to focus on its effect on principal-school matching, which my data is better suited to explore.

[10] Increasingly, districts create standing "talent pools" of teachers and administrators interested in principal positions, but this practice was not used during my sample period.

Consider M schools and $N > M$ potential principals. Student test scores are a function of student ability $\eta_i \sim N(m_s, 1)$ and principal quality $\mu_p \sim U[-\frac{1}{2}, \frac{1}{2}]$. The distribution of student ability is governed by m_s, where for simplicity I assume that a proportion $m_s = 1$ at advantaged schools and $m_s = 0$ at disadvantaged schools. Let $\gamma > 1/2$ be the proportion of schools that are disadvantaged.[11]

Prior to NCLB, schools care about expected test scores minus wages, which due to the rigidity of state salary schedules in North Carolina, are assumed to be fixed. I provide evidence for this assumption in Appendix Table A.

$$\begin{aligned} V_{ps} &= E[\eta_i + \mu_p] - w \\ &= m_s + \mu_p - w \end{aligned}$$

Principal utility is given by:

$$U_{ps} = w + \theta_p m_s + \xi_{ps} \tag{1}$$

where, independent of ability μ_p, principals have preferences $\theta_p \sim N(0, 1)$ over the type of school at which they work. θ_p can reflect a variety of preferences. Some principals may prefer working with disadvantaged students out of redistributive preferences. Alternatively, if succeeding at disadvantaged schools sends a stronger signal of quality, then principals with stronger desires to advance in the career ladder may have stronger preferences for low-performing schools. ξ_{ps} is an infinitesimal idiosyncratic preference, which ensures that principals have strict preferences over schools, but which does not affect anything else.

Proposition 3.1 *There is a unique, stable allocation of principals to schools. Under this allocation, the highest ability principal is matched with his first choice school, the second highest ability principal is matched with her top choice among the remaining vacancies, and so forth until all vacancies are filled.*[12]

Principals with μ_p greater than some threshold μ_A will receive offers from both types of

[11] Assuming $\gamma > 1/2$ merely says that advantaged schools are more scarce and is done to reduce the number of cases. The results of the model would still obtain if the opposite were true.

[12] See Appendix A for proof. I have assumed that schools observe μ_p, but my results are the same if instead principals are ranked by $E[\mu_p]$. My measure of principal ability is informative of the effect of NCLB on mobility as long as it is related to school's perceptions of principal ability.

schools and have the option of working at either. Half the principals, those with $\theta_p > 0$, will choose advantaged schools and the other half will choose disadvantaged schools. Assuming $\gamma > \frac{1}{2}$ so that advantaged schools are scarce, μ_A is determined when advantaged schools fill their vacancies:

$$\frac{N}{2}\left(\frac{1}{2} - \mu_A\right) = (1-\gamma)M \tag{2}$$

This yields

$$\mu_A = \frac{1}{2} - \frac{M}{N}2(1-\gamma).$$

Average quality at advantaged schools is then given by:

$$\begin{aligned} Q_A &= \frac{\frac{1}{2} + \mu_A}{2} \\ &= \frac{1}{2} - \frac{M}{N}(1-\gamma) \end{aligned}$$

Disadvantaged schools fill $(1-\gamma)M$ vacancies with principals who choose to work at disadvantaged schools even though they receive other offers. These are the principals with $\theta_p < 0$ and $\mu_p > \mu_A$. Once these vacancies are filled, there are $\gamma M - (1-\gamma)M = (2\gamma - 1)M$ vacancies remaining. Since advantaged schools have filled all their slots, disadvantaged schools fill these vacancies with principals of quality $\mu \in [\mu_A, \mu_B]$, regardless of their preferences. μ_B solves

$$N(\mu_A - \mu_B) = (2\gamma - 1)M$$

or

$$\mu_B = \frac{1}{2} - \frac{M}{N}.$$

Quality at disadvantaged schools is then a weighted average:

$$\begin{aligned} Q_B &= \frac{1}{\gamma}\left[(1-\gamma)\left(\frac{\frac{1}{2} + \mu_A}{2}\right) + (2\gamma - 1)\left(\frac{\mu_A + \mu_B}{2}\right)\right] \\ &= \frac{1}{2} - \frac{M}{N}\left(2 - \gamma - \frac{1}{2\gamma}\right) \end{aligned}$$

Disadvantaged schools have lower average quality only because advantaged schools are assumed to be scarce: for $\gamma = \frac{1}{2}$, $Q_A = Q_B$. The initial allocation of principals across $\theta - \mu$ space is illustrated in Figure 1. Advantaged schools are filled entirely by principals with $\mu_p > \mu_A$ and $\theta_p > 0$. Disadvantaged schools are filled by principals with $\mu_p > \mu_A$ and $\theta_p < 0$ as well as by any principal with $\mu_p \in [\mu_B, \mu_A)$, regardless of preferences.

Accountability introduces a sanction that principals and schools pay if average test scores fall below a threshold, which I normalize to zero. Principal quality is assumed to affect the test scores of students. A student of ability η_i exposed to a principal of ability θ_p will post a test score of $\eta_i + \theta_p$. In this case, post-accountability principal utility is given by:

$$\begin{aligned} U_{ps} &= w + \theta_p m_s - c \Pr(\eta_i + \mu_p < 0) + \xi_{ps} \\ &= w + \theta_p m_s - c\Phi\left(-\mu_p - m_s\right) + \xi_{ps} \end{aligned}$$

where c is a sanction that a principal pays and Φ is the normal cdf. Similarly, school utility is given by:

$$V_{ps} = \mu_p + m_s - \tilde{c}\Phi\left(-\mu_p - m_s\right) - w.$$

Because of the threshold nature of accountability, disadvantaged schools now value principal quality more than advantaged schools even though there are no complementarities between principal and school quality in the production of test scores. Taking NCLB's stated goals of increasing minimal competency seriously, it is efficient to allocate better principals to disadvantaged schools where they make a greater contribution toward achieving proficiency.

Sanctions associated with student performance, however, make disadvantaged schools relatively less attractive for all principals. Prior to accountability, principals with $\theta < 0$ preferred disadvantaged schools, but afterward, this threshold is pushed to $\theta < g(\mu_p) < 0$ where $g(\mu_p) = -c[\Phi(-\mu_p) - \Phi(-\mu_p - 1)]$ is the difference in expected sanctions between advantaged and disadvantaged schools. $g(\mu_p)$ is always negative but it is increasing in principal ability; the better a principal, the less she worries about being exposed to sanctions. Principals with $\theta_p \in (g(\mu_p), 0)$ change their preference from disadvantaged schools to advantaged schools because their concerns about sanctions outweigh their devotion to working at disadvantaged schools.

Now, when advantaged schools make an offer to a principal that disadvantaged schools also want, they will expect $1 - \Phi(g(\mu_p)) > 1/2$ of them to accept the offer. Because accountability increases yield, vacancies at advantaged schools fill up faster so that only principals of quality μ'_A receive offers from both types of schools. μ'_A solves

$$N\left[1 - \Phi(g(\mu))\right]\left(\frac{1}{2} - \mu'_A\right) = (1-\gamma)M$$

yielding

$$\mu'_A = \frac{1}{2} - \frac{M}{N}\frac{1-\gamma}{1-\Phi(g(\mu_p))}.$$

Since $1 - \Phi(g(\mu_p)) > 1/2$, we can see that $\mu'_A > \mu_A$. Average quality at advantaged schools is given by:

$$Q'_A = \frac{\frac{1}{2} + \mu'_A}{2} = \frac{1}{2} - \frac{M}{2N}\frac{1-\gamma}{1-\Phi(g(\mu_p))} > Q_A$$

Disadvantaged schools receive a lower yield of $\Phi(g(\mu_p))$ so that when advantaged schools have filled up $(1-\gamma)M$ slots, disadvantaged schools have only filled in $N\Phi(g(\mu_p))\left(\frac{1}{2} - \mu'_A\right) = N\left(\frac{1}{2} - \mu'_A\right) - (1-\gamma)M$. Substituting for u'_A, this leaves

$$\gamma M - \frac{\Phi(g(\mu_p))M(1-\gamma)}{1-\Phi(g(\mu_p))} = \frac{M\left[\gamma - \Phi(g(\mu_p))\right]}{1-\Phi(g(\mu_p))}$$

vacancies remaining. These vacancies are filled by principals with quality in (u'_B, u'_A), regardless of preferences:

$$N\left(\mu'_A - \mu'_B\right) = \frac{M\left[\gamma - \Phi(g(\mu_p))\right]}{1-\Phi(g(\mu_p))}.$$

Solving for u'_B yields $u'_B = \frac{1}{2} - \frac{M}{N} = u_B$. This makes sense because the total number of vacancies has not shifted.

Average quality at disadvantaged schools becomes:

$$Q'_B = \frac{1}{\gamma}\left[\left(\frac{\Phi(g(\mu_p))(1-\gamma)}{1-\Phi(g(\mu_p))}\right)\left(\frac{\frac{1}{2}+\mu'_A}{2}\right) + \left(\frac{\left[\gamma - \Phi(g(\mu_p))\right]}{1-\Phi(g(\mu_p))}\right)\left(\frac{\mu'_A + \mu_B}{2}\right)\right] \quad (3)$$

$$= \frac{1}{2} - \frac{M}{N}\left(\frac{2 - \gamma - \frac{\Phi(g(\mu_p))}{\gamma}}{2\left[1-\Phi(g(\mu_p))\right]}\right) \quad (4)$$

When $\Phi(g(\mu_p)) = 1/2$, e.g. when accountability does not diminish the yield for disadvantaged schools, $Q'_B = Q_B$; for lower yields, $Q'_B < Q_B$.

Figure 2 illustrates the shifting distribution of principals across schools after accountability. Disadvantaged schools retain two types of principals after accountability: those with both strong preferences and high-ability who are not deterred by the threat of sanctions ($\mu_p > \mu'_A, \theta_p < g(\mu_p)$), and those who cannot find jobs elsewhere ($\mu_B < \mu_p < \mu'_A$). Principals with $\mu_p > \mu'_A, g(\mu_p) < \theta_p < 0$ are the principals that switch as a result of accountability. Equation (3) is a weighted average of the quality of these groups and captures the intuition that disadvantaged schools are often staffed by a small number of dedicated, high-quality leaders and many more with few other options.

This model makes the following testable predictions:

1. Average principal quality (or perceived quality) declines at disadvantaged schools following the introduction of NCLB.

2. Average principal quality (or perceived quality) increases at advantaged schools following the introduction of NCLB.

3. These effects are greater at schools for which institutionalized sanctions, c and $\tilde{c}$, are greater.

The model does not make an unambiguous prediction about whether high ability principals are more likely to migrate. On the one hand, only principals with quality above u'_A will have the option of moving from disadvantaged to advantaged schools post-accountability. Intuitively, the highest quality principals at disadvantaged schools may not move because they are not worried about sanctions; in this model, however, a subset of them always do because they do not have strong preferences (θ_p negative, but near zero) that would compel them to stay. On the other, this model also predicts that there will be movement among lower quality principals who used to work at advantaged schools but are now forced out as a result of the influx of higher quality principals formerly at disadvantaged schools. Which effect dominates remains an empirical question.

The conceptual framework presented above differs from actual principal-school matching in several ways. First, I have implicitly assumed that principals can be displaced. In reality, this is unlikely to be the case, so that NCLB's impact on mobility and quality is bounded by the number of vacancies. More generally, differences in queues ex ante at different schools affect the extent to which changes in principal preferences translate into assignment of principals to schools. Schools with long queues are less likely to see a change in the average quality of their principals because marginal changes to the applicant pool are less likely to make a difference in terms of who is hired. Engel and Jacob (2011) show that teachers in the Chicago Public School system are more likely to show interest in schools with lower poverty rates. The quality of principals at Title I schools may be more sensitive to accountability pressures both because sanctions are stronger and, potentially, because queues may be shorter. The predictions of this model are also bound by the number of school's in a principal's choice set. Thus, effects may be also be stronger for large and urban districts where principals have a larger choice set of schools. Results by district characteristics are reported in Table 6.

4 Empirical Methods

I test the predictions of the model in Section 3 by providing estimates of principal quality μ_p based on principal performance in the period prior to the implementation of NCLB. I then identify schools that are likely to fail AYP based on an index of student demographics from 1995 to 2002. Combining these measures of principal and school performance, I examine the effect of NCLB's threat of sanctions on the distribution of principal quality across schools and on principal mobility. I check if larger sanctions c lead to larger declines in principal quality by examining the effect of NCLB on Title I schools, which are subject to official AYP sanctions compared with non-Title I schools, which are not.

Principal quality is difficult to estimate because it requires separating the effect of a principal on student achievement from unobserved neighborhood or school effects. A principal in one school may have advantages over a principal in another that cannot be captured by controls for school budgets or demographics alone: parental motivation, supportive school boards, and local supplies of teachers are all factors that are difficult to control for, but which may substantially impact student performance.

I quantify principal quality using the following model decomposing student performance into individual, school, and principal components using variation from principal mobility across schools:

$$y_{ispt} = \beta_0 y_{isp't} + \beta_1 X_i + \beta_2 X_{st} + \mu_s + \mu_p + \mu_{t\times g} + \varepsilon_{ispt}. \tag{5}$$

Here, y_{ispt} is an outcome for student i at school s in year t under principal p, X_{it} are student demographics, X_{st} are time-varying school characteristics, μ_s are school fixed effects, μ_p are principal fixed effects, $\mu_{t\times g}$ are year-grade fixed effects, and ε_{ispt} is an error term. Typically, teacher value-added regressions include controls for lagged scores, but in the case of principals, doing so ignores the cumulative effects of principals over multiple years. Instead, I include controls $y_{isp't}$ for the most recent test score under previous principals p', if available. The inclusion of school fixed effects controls for persistent differences in student and staff quality. School fixed effects and lagged scores for potentially non-random sorting by principals into schools. Year by grade fixed effects control for time-varying differences in testing regimes.

The variance of the measured fixed effects $\hat{\mu}_p$ in Equation (5) overstates true variance in principal quality because it reflects both variation in true principal quality and measurement error. Following the spirit of Kane and Staiger (2008), I adjust these estimates using an Empirical Bayes estimator to shrink high variance observations toward the mean: $VA_p = \lambda_p \hat{\mu}_p$, where λ_p is a principal specific shrinkage factor. Details are described in the Appendix.

The principal fixed effects in Equation (5) cannot be identified for principals who stay at a school for the entire duration of the sample period because their contribution cannot be distinguished from a school fixed effect. The remaining principals for whom fixed effects can be identified include principals who are only observed in one school, and who stay for a proper subset of the sample period (newcomers or leavers), and those who are observed at multiple schools (switchers). Fixed effects for non-switchers are confounded with time-school-specific effects that may plausibly be attributed to a host of unobservable factors. As such, I focus on switchers only and attribute principal effectiveness to the portion of student achievement that is correlated across schools that a principal is observed in, but which is not explained by other observables and fixed effects.

Identifying principal effects from movers mitigates concerns about conflating school and principal effects, but introduces new selection issues. Principals are not randomly assigned to schools, and if principals systematically move based on the achievement *gains* of students, then the fixed effects estimated in Equation (5) may conflate other reasons for changes in performance with true principal effects. Rothstein (2007) shows, in the context of estimating teacher value-added, that test score gains of students can be predicted by the value-added of their future teachers, indicating that teachers are being assigned to classrooms based on student test score gains. Since scores tend to be mean-reverting, a teacher who is assigned to students with high gains in the previous year is unfairly penalized when score gains likely decrease in the current year.

Rothstein's concerns, however, are less of a problem in the context of studying principals. While principals have substantial knowledge about the test scores and other characteristics of students in their own school and may use this information in assigning teachers to classrooms, they have less information about the test score gains of students at other schools and are thus less likely to use this information in their own mobility decisions.

Another concern about principal quality is that it may evolve over time. If much of a principal's true effectiveness comes from learning, this is not reflected in the fixed effect. Instead of including principal fixed effects in (5), I could have included principal covariates such as tenure, experience, and education. Previous research on both teachers and principals, however, indicates that the vast majority of variation in educator quality cannot be explained by observables.[13] Thus, I use principal value-added as an imperfect measure of full variation in principal ability.

More generally, this study is concerned about principal quality insofar as it informs the allocative effects of NCLB. As a result, potential bias in value-added is less problematic for three reasons: first, estimates of changes in the assignment of principals to schools based on value-added reflect changes in the true distribution of quality as long as the bias in principal value-added is systematic across principals; second, value-added may be reflective of perceived principal quality and thus be nonetheless informative about the labor market opportunity of principals; and third, mismeasurement of either perceived or true quality biases me away from finding a systematic relationship between mobility and quality as a result of NCLB. It is worth emphasizing the third point here; if Equation (5) produced estimates of principal value added that are not reflective of either principal quality or perceived principal quality, then we do not expect the distribution of this measure across high- and low-performing schools to systematically change after the implementation of NCLB.

[13] See Kane, Rockoff, and Staiger (2007) for teachers and Branch, Hanushek, and Rivkin (2012) for principals.

Using these estimates of principal quality, I next estimate the impact of NCLB on the allocation of principal quality across schools. To conduct this analysis, I exploit exogenous variation in the likelihood, $P = \Phi(-\mu_p - m_s)$, that a principal faces performance sanctions arising from variation in m_s, the baseline ability of students in school s. Schools with low m_s, for instance those serving disadvantaged student populations, have a higher likelihood of facing sanctions, independent of a principal's ability or actions. I quantify the portion of P that is due to m_s alone by estimating the probability that a school fails AYP based on student demographics only. This characterizes a school's probability of failure for which a principal should not, in theory, be penalized.

Under NCLB, a school passed AYP if it met performance targets for *all* of its qualifying demographic subgroups. This means that demographic variation in the likelihood that a school fails AYP comes from two main sources. First, schools with larger proportions of historically poorly performing students were more likely to fail, because performance targets for disadvantaged groups were set at the same level as those for advantaged groups. Second, schools were larger numbers of minority groups were more likely to fail, simply because they faced a larger number of performance targets to meet (a school failed AYP if it failed in any subgroup). Specifically, a subgroup's performance does not count for AYP if it there are fewer than 40 members of that group. I estimate the probability that a school fails AYP based on pre-period demographics only, taking into account both these sources of variation:

$$\Phi(\text{fail}_s) = X_s\beta + \varepsilon_s. \tag{6}$$

In Equation (6), fail_s is an indicator for whether school s would fail AYP in 2001-2002 under 2002-2003 rules, based on the number of demographic subgroups in the school, their performance, and the size of those subgroups. I then use a school's demographics prior to 2002 to predict this measure of performance. Because I am not predicting actual AYP status, which could be influenced by the implementation of NCLB, I treat $\Phi(\text{fail}_s)$ as a known demographic index that describes principals' perceptions of their likelihood of failure in 2002.

The covariates X_s include, for each year from 1995 to 2002, cubics for racial composition, proportion of students eligible for free lunch, percentage of students with a parent with some post-secondary education, and school size, with linear effects that are allowed to be different for K-5 schools and non K-5 schools, and the number of students in each demographic subgroup. I also include dummies for whether each of the NCLB-defined subgroups were large enough to be counted for AYP, in both reading and math.[14] This specification allows both for the proportion of students who belong to a subgroup to impact a school's probability of failure, as well as for the size of these subgroups to matter. X_s also includes dummies for whether a school is K-5 or urban. The fitted probability of failing becomes my measure of the inherent likelihood of facing sanctions for principals working at each school.

$\Phi(\text{fail}_s)$ indexes a school's exposure to NCLB sanctions and is fixed across schools over time. In reality, however, probabilities of failure change for a school over time either due to changes in student performance or changes in target thresholds so that $\Phi(\text{fail}_s)$ may not necessarily reflect the likelihood of failure for later years in the post-NCLB period. Constructing my measure of failure probability to reflect real probabilities of failure, however, produces a measure of exposure that is endogenous to principal performance. I choose a static measure of likelihood of failure in order to capture the part of NCLB risk that is outside of a principal's control.

Restricting to principals for whom I have estimated pre-period quality and extending the sample period to follow those principals in the post-NCLB years, I ask whether principal quality at disadvantaged schools changes relative to advantaged schools following the implementation of NCLB. The estimating equation is given by:

$$\begin{aligned} VA_{pst} &= \alpha_0 + \alpha_1 \Pr(\text{fail})_s \times \mathbb{I}\{\text{year} > 2002\} + \alpha_3 \Pr(\text{fail})_s \\ &\quad + \alpha_4 \mathbb{I}\{\text{year} > 2002\} + X_{pst} + \delta_d + \delta_t + t \times \delta_d + \varepsilon_{pst} \end{aligned} \tag{7}$$

where VA_p is estimated principal quality, Pr(fail) is a school's probability of failing AYP, X_{pst} are principal covariates, and δ_d, δ_t, and $t \times \delta_d$ are, respectively, district and year fixed effects, and district linear time trends. District specific time trends allow principal quality among high- and low-performing schools to be on different trends across districts. The possibility that districts are on separate trends is particularly likely in North Carolina, which includes both rural and urban districts with significant variation in racial composition. In this specification, α_1 identifies the

[14] For this calculation, there are 16 targets: math and reading targets for Black, White, Hispanic, Asian, Native American, male, female, and all students.

effect of NCLB under the assumption that, within districts, high- and low-performing schools are on stable trends in the absence of NCLB.

Using a complementary specification, I also estimate the effect of NCLB on measures of turnover by substituting mobility variables in the left hand side of (7) and examining the impact of NCLB on the characteristics of the next school to which principals are assigned. The estimates of NCLB's effect on aggregate principal mobility can be further refined to investigate heterogeneity in principal mobility by ability. I allow the effect of NCLB to differ for principals above and below median estimated quality and test whether high-ability principals are more likely to move, and, conditional on moving schools, what are the characteristics of their new schools.

5 Data

I use administrative records from the North Carolina Public School System. These data have been compiled by the North Carolina Education Research Center into student-school and staff-school matched panels spanning the years 1995 through 2007. These data include a unique staff ID that allows me to track principals as long as they move within the state.

5.1 Sample Construction

I estimate Equation (5) using student level data in the period prior to NCLB, from 1995 to 2002. Figure 3 outlines my procedure for constructing the final analytic sample. From an initial sample of 4,890 full-time principals in schools which employ at most one principal at a time from 1995 to 2002, I match on student test scores and restrict to student-year observations for which 1) I have data on both math and reading test scores for the current and previous year, 2) schools where there are at least two observed principals in the pre-period, and 3) schools where at least one principal is a switcher in the pre-period.[15] For each of these schools, I retain all observations, including those for years in which the school principal is not a switcher. This yields a subsample of 500 schools and 832 principals. In estimating principal fixed effects, I specify that all school fixed effects must be estimated; this allows me to estimate principal fixed effects for 640 principals, of whom 298 are

[15] Not all years are represented in this dataset because test scores are available only for a subset of years and grades, so that, strictly, a principal must move from one school-year with test scores to another school-year with test scores before 2002.

movers.[16]

To study compositional effects, I follow these principals in the post-NCLB years. This initially expands the number of schools in my sample to 596, but I restrict the sample to schools with standard grades that remain open for the entire sample period from 1995-2007 to avoid spurious mobility effects coming from school openings and closings. Approximately a third of schools are not observed in all years. The final analytic sample includes observations on 214 principals in 383 schools. Each school is observed for an average of six years over the period 1995 to 2007.

5.2 Descriptive Statistics

Estimating principal quality from the subset of principals that switch schools prior to the implementation of NCLB creates a measure of quality that is less likely to be contaminated by unobserved school effects. The cost, however, is that this sample of switcher principals may systematically differ in a way that limits the external validity of my estimates.

Table 1 shows summary characteristics of principals and schools in the analytic sample compared to the universe of principals who are in the school system prior to NCLB. There are significant differences between the two samples. Sample principals are observed in my data for approximately 1.5 years more. By construction, all of them have switched schools at least once in my sample period, compared to 66 percent for the universe of principals. Both sets of principals appear to switch at the same time in their careers, early on in their first principalship while they are still under provisional contracts. The schools represented in my analytic sample are slightly more likely to fail AYP. Sample schools are also more urban, have slightly higher minority shares, are more likely to receive Title I funds, and are much more likely to be K-5 elementary schools. Principal salary and tenure are similar. These differences are logical since principals of elementary schools and those working in urban districts may plausibly have more nearby employment options and thus be more likely to have switched schools once in the past.

School characteristics are likely to be correlated with principal mobility patterns (e.g. Cullen and Mazzeo, 2007). My study, however, would have lower external validity if school characteristics were *differentially* correlated with mobility patterns, compared with the universe of principals.

[16] In a school with two principals, only fixed effects for one principal can be estimated if school fixed effects are also included. The final principal serves as a reference.

Table 2 shows that this is not the case. Each panel of Table 2 asks whether sample principals are representative of the universe in terms of how their mobility is correlated with a particular school characteristic: its probability of failing AYP based on demographics, the proportion of students who are white, and the proportion of students who are eligible for free or reduced price lunch. I show that while sample principals are more likely to switch schools by construction, they do not differentially prefer to leave certain types of schools as compared with the universe of principals. For example, Table 2 Panel 2 shows that sample principals are no more or less likely to switch out of a school on the basis of the proportion of white students than the universe of principals. Other issues of sample selection are discussed in the Appendix.

Table 3 reports coefficients from a regression of a school's probability of failure measure defined in Equation (6) on a selected set of school characteristics (this is not the actual equation used to estimate Pr(Fail), which involves demographic subgroup sizes interacted with school level and other variables). The excluded categories are white students and those not eligible for free lunch, so that the coefficients in Table 3 are of the expected sign: a school is at greater risk of facing AYP sanctions if it serves more minority and low income students.

6 Results

6.1 Principal Quality Estimates

I first estimate principal fixed effects from Equation (5) and then adjust for measurement error. There is substantial variation in principal quality. Figure 4 plots the estimated distribution of principal quality in math and reading. The dashed line represents principal quality before applying the shrinkage procedure discussed in Section 4. The shrinkage estimator compresses estimates of principal quality and has the greatest effect at the tails of the principal quality distribution. Nonetheless, principal quality, even adjusted for measurement error, remains highly variable: a one standard deviation increase in principal math quality is predicted to increase the math test score of an average student by a fifth of a standard deviation relative to other North Carolina students in that grade and year. These effects are about twice as large as those estimated for teachers, but come from the fact that I allow for principal effects to accumulate over multiple years by only controlling for prior test scores under a different principal. Principal reading quality is closely correlated with math quality (correlation: 0.727), but variation in reading effects is smaller.

Principal quality varies systematically with school quality as measured by a school's probability of failing AYP. Figure 5 plots the distribution of estimated principal quality in math for schools above and below the median probability of failure before and after the implementation of NCLB. At low-performing schools (Figure 5, top panel), the lower tail of principal math quality shifts further down after 2003, whereas at high-performing schools, the distribution of principal quality improves slightly after NCLB (Figure 5, bottom panel). The distribution of principal quality in reading (Figure 6) follows a similar pattern: the distribution of quality at low-performing schools remains similar or shifts slightly lower, while the distribution of principal quality at high-performing schools significantly improves following NCLB.

Figures 5 and 6 also report estimates of the variation in principal value-added, adjusted for measurement error. I find that principal value-added ranges from 0.18 to 0.27 in math and 0.14 to 0.19 in reading. These numbers mean that a one standard deviation increase in principal quality is predicted to increase student test scores by about one-fifth of a standard deviation. I also find that principal quality is more variable in low-performing schools than in high-performing ones and that this difference increases after NCLB. Both of these results are consistent with my model, which predicts that good principals will leave low-performing schools in order to avoid sanctions, but that the best principals will stay because they are confident in their abilities. My estimates of variation in value-added are also consistent with Branch, Hanushek, and Rivkin (2012) whose estimates, based on Texas data, range from 0.20 to 0.24.

These differences in the distribution of principal quality translate into economically significant differences in the access that various demographic groups have to a high-quality principal. Each cell in Table 4 reports results from a regression of estimated principal quality on school characteristics, controlling only for district fixed effects. I include district fixed effects because principals typically move within the same district, so that these results are informative about the correlation between principal quality and school demographics among schools in the same district, which are more likely to be in a particular principal's choice set. The results in Table 4 indicate that prior to the implementation of NCLB, principal quality is correlated with student performance variables, but not significantly correlated with student demographics like race. After NCLB, however, students from disadvantaged backgrounds become significantly less likely to attend a school with a high-quality principal. The magnitude of this change in correlation is sizable; in the pre-period, the percent of black students is not correlated with principal math quality, controlling for overall student performance. After NCLB, a 10 percentage point increase in the percent of black students is correlated with a -0.306*(0.10)=-0.031 point decrease in estimated principal quality. This works out to a -0.031/0.232=-0.134th of a standard deviation decrease in quality.

This result is suggestive of an adverse allocative effect of NCLB: by defining AYP in terms of thresholds that are more difficult to meet at schools with more students from disadvantaged backgrounds, NCLB effectively penalizes principals for the demographics of their students. Table 4 indicates that high-quality principals seem to respond to these incentives by choosing to work at schools with fewer disadvantaged students.

6.2 Impact of NCLB on Quality

I next examine the effect of NCLB on the allocation of principal quality across schools in more detail. Using estimated principal quality as outcomes in Equation (7), I find that NCLB leads to systematic declines in principal math quality at schools whose demographics put them at greater risk for failing AYP (Table 5, Column 1). This effect, moreover, is driven by falling principal quality at Title I schools, which are directly subject to AYP sanctions (Table 5, Column 2). This result is consistent with the model in Section 3, which predicts that quality effects are smaller when the probability of facing sanctions is lower.

The magnitude of the effect of NCLB on principal quality is sizable. The results in Column 2 of Table 5 show that at Title I schools a one standard deviation higher likelihood that a school fails AYP (0.417) leads to a $0.417 \times 0.228 = 0.095$ point decline in math effectiveness attributable to NCLB. Given that the standard deviation of principal math ability is 0.232, this translates into a decline in principal math ability of one third to one half of a standard deviation. Recalling that a one standard deviation higher-ability is associated with a fifth of a standard deviation increase in test scores, this means that given two Title I schools one standard deviation apart in failure probability, students at the worse school are expected to lose approximately 9.5% of a standard deviation in math test scores as a result of NCLB.

Most studies focus on the effect of NCLB on principal and teacher *effort* and, generally, find zero to positive effects on test scores.[17] The logic is that principals and teachers respond to accountability by working harder (or engaging in more gaming behaviors). My results, however, capture a different dimension of how NCLB can affect student test scores. Because I measure principal quality using test-scores in the *pre*-period, I, by design, cannot capture the effect of NCLB on principal effort or gaming. Instead, my negative estimates of the effect of NCLB on principal quality (and thus test scores) capture the *allocative* effect of NCLB on the quality of school leaders. The logic here is that NCLB can cause student test scores at disadvantaged schools to fall relative to scores at more advantaged schools by encouraging high-performing staff to work elsewhere.

I find that principal quality in both math and reading fall at Title I schools, but not at non-Title I schools. First, non-Title I schools are not directly subject to NCLB sanctions so it is unsurprising that their principal quality is less affected. Second, I will provide evidence in Table 8 that declines in principal quality at disadvantaged non-Title I schools were attenuated by increased migration of high value-added principals from disadvantaged Title I schools.

Table 6 examines heterogeneity of the effect of NCLB across districts. I find that the effects of NCLB are stronger in districts where principals are likely to have more mobility—ones with more schools or those in urban areas where schools are closer.[18] This is consistent with the model in which principal mobility drives changes in the distribution of quality.

The regressions in Tables 5 and 6 control for district and year fixed effects, principal age and age squared, and a linear district time trend. I include district fixed effects and district by year time trends to allow for the possibility that principal quality among high- and low-performing schools may be on different trends, depending on the district. Given that North Carolina includes both rural and urban districts with significant variation in racial composition, separate trends by district may be likely.

Next I examine the timing of the decline in principal quality. If the decline in principal quality at Title I schools documented in Table 5 occurs as a result of NCLB, relative principal quality should not depart from its trend until after NCLB is signed into law in 2002 or implemented in 2003. Figure 7 plots the effect of NCLB for each year to show that this is indeed the case. For both principal

[17] Figlio and Rouse (2006), West and Peterson (2006), Rouse et. al. (2007), Chiang (2008), Krieg (2008), Neal and Schanzenbach (2007), and Dee and Jacob (2009) all find test score gains of some kind.

[18] These results are not driven solely by Charlotte-Mecklenburg and hold up to its exclusion from the sample.

quality in reading and math at Title I schools, there is a break in pre-period trends around the time that NCLB is implemented, consistent with a causal impact of NCLB on the distribution of principal quality. The effect of NCLB on principal quality increases with time, flattening out at the end of the sample period. This is consistent with the fact that it may take a few years for principals to move, especially because most moves occur at the end of a principal's four-year contract period.

Principal quality in non-Title I schools does not seem to be affected; this may be because they are not directly subject to NCLB sanctions or because high quality principals begin to differentially prefer non Title I schools following NCLB (Table 8, Column 5). The results in Table 5 and Figure 7 indicate that the implementation of NCLB lead to a decline in principal quality at schools most likely to fail AYP based on their demographics.

6.3 Impact of NCLB on Mobility

In this section, I analyze possible mechanisms underlying the decline in average principal quality at disadvantaged schools. Table 7 reports estimates of Equation (7) where the outcome of interest is a dummy for whether a principal moves to a different school in the next year. I do not find evidence that NCLB increases the aggregate likelihood that principals of high-risk schools switch jobs, either in my analytic sample or in the universe of principals. The final column of Table 7 examines the impact of NCLB on retirement rates for the universe of principals. I do not report estimates for the sample of principals for whom I have estimates of quality because my sample construction method—requiring that principals switch schools at least once during the pre-period—yields artificially low retirement rates before 2003. Column 3 indicates that principals do not seem to differentially retire more at schools more likely to fail AYP.

The allocative impact of NCLB as described in my model, however, comes from heterogeneity in principal mobility patterns by ability. Despite no aggregate effect, Table 8 shows that principal ability is indeed linked to subsequent mobility choices; the threat of sanction appears to affect where principals choose to work next, conditional on switching schools. I regress the qualities of the next school a principal works at on the extent to which a principal's current school is affected by NCLB. Specifically, Table 8 reports coefficients on $\Pr(\text{fail}) \times \mathbb{I}\{\text{year} > 2002\}$ for principals above and below median math quality.

Columns 2 through 5 of Table 8 indicate that NCLB leads higher-ability principals at poorly performing schools to, conditional on switching, move to schools with lower probabilities of failure, more students at grade level, a larger proportion of white students, and to non-Title I schools.[19] Consider two effective principals who, prior to NCLB, serve at an advantaged and disadvantaged school, respectively. These results say that, after NCLB is implemented, the principal serving at the disadvantaged school is *differentially* more likely to move to a higher-performing school than the principal serving at the advantaged school, as compared to before NCLB. Because this estimate is a difference-in-difference, this finding cannot be attributed to extant patterns of career progression as described in Mazzeo and Cullen (2007).

One potential benefit of accountability is that increased scrutiny may increase retirement among low-skill principals. Here, my sample selection criterion prevents me from estimating retirement effects, since the sample requirement that principals appear at least once in the post-period mechanically restricts my sample to principals who do not retire in the pre-period. Thus I cannot estimate baseline retirement rates in the pre-period for the sample of principals with quality estimates.

6.4 Accountability under Growth vs. Levels: NCLB and the ABCs of Growth

The transition from North Carolina's existing accountability system, the ABCs of Growth, to new NCLB performance standards disproportionately affected high minority-share schools with low test scores. Because the ABCs based targets on growth in scores while NCLB focused primarily on level targets, the schools that experienced the greatest shock were those that were improving test scores among poorly performing students; in fact, almost 20% of schools which failed more than 25 percent of their subgroup-specific AYP targets in the first year of NCLB received the highest ABCs rating, "Exemplary," the prior year.

Table 9 considers whether NCLB had a particularly detrimental effect on schools that were formerly meeting growth targets under the ABCs. Principals which were formerly meeting or exceeding standards under ABCs, but who then failed to meet AYP requirements may have experienced a greater "shock" to job pressure and security. I define shocked schools as those which met ABC requirements in 2002, but which failed at least 25 percent of their AYP subgroup targets in their first year under NCLB and redo my analysis in Tables 5 and 8 with this as the treatment

[19]These results hold when principals are split into terciles or quartiles as well.

variable.[20] This presents an alternate way of identifying the schools most likely to be affected by NCLB targets.

In my data, schools that experienced a large shock with the introduction of NCLB had fewer students performing at grade level (69% vs. 77%) and more minority students (34% white vs. 58%). This is consistent with the fact that schools with low test scores could meet ABC requirements by improving test scores while it was very hard to do the same thing under NCLB (starting in 2003-04, schools could make AYP through NCLB's Safe Harbor provisions if it improved test scores by over 10% for all below-target demographic subgroups. In practice, however, few schools did this–see Section 2 for more details).

Column 1 of Table 9 indicates that NCLB led principal math quality at shocked schools to fall by 0.175/0.232 or 0.754ths of a standard deviation, relative to non-shocked schools. Similar to the main results in Table 5, this effect is driven primarily by quality falling at Title I schools (Column 2). NCLB also does not appear to cause principals at shocked schools to differentially move (Column 4), but does appear to influence the qualities of the next school a principal works at, conditional on moving. In particular, better principals appear to disproportionately leave shocked schools for schools with higher *levels* of student achievement.

7 Conclusion

Much research on NCLB has focused on school and teacher efforts to increase student test scores. School staff, however, can respond to the pressures of accountability not only by altering the types of effort it puts toward improving test scores, but also by choosing where to work. A primarily contribution of this paper is to quantify the importance of this mobility response for the allocation of qualified staff across high- and low-performing schools. To do this, I analyze the impact of NCLB on the labor market for school principals. I develop a theoretical framework highlighting the consequences of an uncompensated change in the likelihood that a principal faces sanctions on his or her subsequent mobility and test this model using the implementation of NCLB. I find that NCLB leads to declines in the math and, to a lesser extent, reading quality of principals assigned to schools more likely to face sanctions. As predicted by my model, I find that declines in principal

[20] This measure of shock is robust to alternative specifications of the percent of AYP targets failed.

ability at disadvantaged schools are caused by the departure of high-quality principals for schools where these principals have a lower likelihood of facing AYP sanctions.

More broadly, this paper shows that in order to evaluate the success of accountability policies such as NCLB, one needs also to consider its impact on incentives in the broader labor market for educators. As a policy that elevates minimal competency to the forefront of educational goals, NCLB demands that greater resources be allocated to students for whom the presence of a high-quality educator may push them pass the proficiency threshold. Implementing NCLB without fully compensating principals for the increased penalties associated with working with disadvantaged student populations, however, leads to the opposite allocative effect. This paper suggests that districts or policy makers may want to consider the effects of NCLB on the distribution of talent across schools when setting wages or evaluating accountability practices.

References

[1] Ballou, Dale. 1996. "Do Public Schools Hire the Best Applicants?" *The Quarterly Journal of Economics*, 111(1): 97-133.

[2] Ballou, Dale, and Michael Podgursky. 1995. "What Makes a Good Principal? How Teachers Assess the Performance of Principals." *Economics of Education Review*, 14(3): 243-252.

[3] Ballou, Dale, and Michael Podgursky. 2002. "Seniority, Wages and Turnover Among Public School Teachers". *Journal of Human Resources*, 37(4): 892-912.

[4] Bertrand, Marianne, and Antoinette Schoar. 2002. "Managing With Style: The Effect Of Managers On Firm Policies." *The Quarterly Journal of Economics*, 118(4): 1169-1208.

[5] Besley, Timothy, and Stephen Machin. 2008. "Are Public Sector CEOs Different? Leadership Wages and Performance in Schools." http://econ.lse.ac.uk/staff/tbesley/papers/pubsecceo.pdf.

[6] Billger, Sherrilyn. 2007. "Principals as Agents? Investigating Accountability in the Compensation and Performance of School Principals." *Industrial and Labor Relations Review*, 61(1): 90-107.

[7] Boyd, Daniel, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. "The Impact of Assessment and Accountability on Teacher Recruitment and Retention: Are There Unintended Consequences?" *Public Finance Review*, 36(1): 88-111.

[8] Branch, Gregory, Eric Hanushek, and Steven Rivkin. 2012. "Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals." NBER Working Paper #17803.

[9] Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics*, 93(9-10): 1045-1057.

[10] Clotfelter, Charles, Helen Ladd, and Jacob Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources*, 41(4): 778-820.

[11] Cullen, Julie and Michael Mazzeo. 2007. "Implicit Performance Awards: An Empirical Analysis of the Labor Market for Public School Administrators." Working Paper.

[12] Dee, Thomas, and Brian Jacob. 2009. "The Impact of No Child Left Behind on Student Achievement." NBER Working Paper #15531.

[13] Eberts, Randall, and Joe Stone. 1988. "Student Achievement in Public Schools: Do Principals Make a Difference?" *Economics of Education Review*, 7(3): 291-299.

[14] Eeckhout, Jan. 2000. "On the uniqueness of stable marriage matchings." *Economic Letters*, 69: 1-8.

[15] Engel, Mimi, and Brian Jacob. 2011. "New Evidence on Teacher Labor Supply." NBER Working Paper #16802.

[16] Figlio, David and Lawrence Getzler. 2002. "Accountability, Ability, and Disability: Gaming the System," NBER Working Paper #9307.

[17] Figlio, David, and Cecilia Rouse. 2006. "Do Accountability and Voucher Threats Improve Poorly-Performing Schools?" *Journal of Public Economics*, 90(1-2): 239-255.

[18] Figlio, David, and Joshua Winicki. 2005. "Food for Thought: the Effects of School Accountability Plans on School Nutrition." *Journal of Public Economics*, 89(2-3): 381-394.

[19] Fryer, Roland, Steven Levitt, John List, and Sally Sadoff. 2012. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment." NBER Working Paper #18237.

[20] Hanushek, Eric, John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. "The Market for Teacher Quality." NBER Working Paper #11154.

[21] Hanushek, Eric, and Margaret Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, 24(2): 297- 327.

[22] Hanushek, Eric, and Steve Rivkin. 2010. "Does Teacher Job Search Harm Disadvantaged Urban Schools?" NBER Working Paper #15816.

[23] Jackson, Kirabo. (2011) "School Competition and Teacher Labor Markets: Evidence from Charter School Entry in North Carolina." NBER Working Paper #17225.

[24] Jacob, Brian. 2005. "Accountability, Incentives, and Behavior: The Impact of High- Stakes Testing in the Chicago Public Schools. " *Journal of Public Economics*, 89 (5-6): 761-796.

[25] Jacob, Brian. 2010. "Do Principals Fire the Worst Teachers?" NBER Working Paper #15715.

[26] Jacob, Brian, and Steven Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *The Quarterly Journal of Economics*, 118 (3): 843-877.

[27] Kane, Thomas, Jonah Rockoff, and Doug Staiger. 2007. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review*, 27(6): 615-631.

[28] Kane, Thomas, and Doug Staiger. 2008 "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper #14607.

[29] Knapp, Michael, Bradley Portin, Michael Copland, and Margaret L. Plecki. 2006. "Leading, Learning, and Leadership Support." Monograph. University of Washington Center for Teaching and Policy and The Wallace Foundation.

[30] Krieg, John. 2008. "Are Students Left Behind? The Distributional Effects of No Child Left Behind." *Education Finance and Policy*, 3(2): 250-281.

[31] Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: a Descriptive Analysis." *Educational Evaluation and Policy Analysis*, 24(1): 3862.

[32] Lavy, Victor. 2008. "Does Raising the Principals Wage Improve the Schools Outcomes? Quasi-experimental Evidence from an Unusual Policy Experiment in Israel." *Scandinavian Journal of Economics*, 110(4): 639-662.

[33] Loeb, Susanna, Demetra Kalogrides, and Eileen Horng. 2010. "Principal Preferences and the Uneven Distribution of Principals Across Schools" *Education Evaluation and Policy Analysis*, 32(2): 205-229

[34] Loeb, Susanna, Demetra Kalogrides, and Tara Beteille. "Effective Schools: Teacher Hiring, Assignment, Development, and Retention." NBER Working Paper #17177.

[35] Neal, Derek, and Diane Schanzenbach. 2007. "Left Behind By Design: Proficiency Counts and Test-Based Accountability." NBER Working Paper #13293.

[36] Reback, Randall. 2008. "Teaching to the Rating: School Accountability and the Distribution of Student Achievement." *Journal of Public Economics*, 92(5-6): 1394-1415.

[37] Reback, Randall, Jonah Rockoff, and Heather Schwartz. 2011. "Under Pressure: Job Security, Resource Allocation, and Productivity in Schools Under NCLB." NBER Working Paper #16745.

[38] Rockoff, Jonah. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review Papers and Proceedings*, 94(2): 247-252.

[39] Rockoff, Jonah, Douglas Staiger, Thomas Kane, and Eric Taylor. 2010. "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools." NBER Working Paper #16240.

[40] Roth, Alvin, and Marilda Sotomayor. 1990. *Two-sided Matching: a Study in Game-theoretic Modeling and Analysis.* New York: Cambridge University Press.

[41] Rothstein, Jesse. 2007. "Do Value-Added Models Add Value? Tracking, Fixed Effects, and Casual Inference." CEPS Working Paper #159.

[42] Rouse, Celia, Jane Hannaway, Dan Goldhaber, and David Figlio. 2007. "Feeling the Florida Heat? How Poorly-performing Schools Respond to Voucher and Accountability Pressure." NBER Working Paper #13681.

[43] West, Martin, and Paul Peterson. 2006. "The Efficacy of Choice Threats within School Accountability Systems: Results from Legislatively Induced Experiments." *Economic Journal*, 116(510): C46-C62.

A Proof of Proposition 3.1

I first show that there is a unique, stable allocation of principals to schools where the principal with the highest μ is matched to his top choice school, and the principal with the second highest μ is matched to her top choice among the remaining schools, etc., until all vacancies are filled. Assume that the pairings (p_i, s_i) result, where i is the rank of the principal, and s_i is the school chosen by principal i in the manner just described.

Suppose, however, that there exists a blocking pair (p_i, s_j) for $i \neq j$. Because schools share rankings, in order for school j to prefer i, it must be that $i > j$. However, p_i prefers s_i to any s_j for $j > i$ because school j was in p_i's choice set. Thus, it could not be the case that (p_i, s_j) is a blocking pair. This shows that the proposed allocation is stable.

Suppose further that there exists any other stable allocation. This means that for some i, p_i is not matched with s_i. If p_i is matched with s_j for $j < i$, then school j prefers principal j to i. In order for (p_j, s_j) to not be a blocking pair, it must be that principal j is matched to a school he prefers to s_j, call it s_k. For this to be true, it must be that $k < j$. Then, in order for (p_k, s_k) to not form a blocking pair, principal k must be matched to some s_n, $n < k$. Continuing in this way, we reach a contradiction that school s_1 is matched to some principal other than p_1.

If, on the other hand, p_i is matched with s_j for $i < j$, principal i prefers school i. Thus (p_i, s_i) is a blocking pair unless s_i is matched with p_k, $k < i$. The same contradiction follows.

B Value-added Adjustment

Principal fixed effects $\hat{\mu}_p$ estimated from Equation (5) include estimation error so that, ignoring potential bias, $\hat{\mu}_p$ is a combination of the true effect plus a noise term I assume to be independent and normal:

$$\hat{\mu}_p = \mu_p^* + \nu_p \tag{8}$$

In this case, $\text{Var}(\hat{\mu}_p) = \text{Var}(\mu_p^*) + \text{Var}(\nu_p)$ so that the estimate of true variance is upwardly biased from additional variance coming from estimation error.[21] To correct for this, I note that the best estimate for μ_p^* is given by $E(\mu_p^* | \hat{\mu}_p) = \lambda \hat{\mu}_p + (1-\lambda)\overline{\hat{\mu}_p}$ where $\overline{\hat{\mu}_p} = 0$ by design and $\lambda_p = \frac{\sigma^2_{\mu^*}}{\sigma^2_{\mu^*} + \sigma^2_\nu}$ is a shrinkage term constructed as the ratio of the estimated variance of true principal effects $\sigma^2_{\mu^*}$ to the sum of estimated true variance $\sigma^2_{\mu^*}$ and estimated noise variance σ^2_ν.

In the teacher effects literature, the common solution to this measurement error problem is to use across-time correlation in estimates of teacher fixed effects to construct λ. Applying this approach to principals, however, requires data on multiple principal moves, which happens rarely in practice: in my sample, only 6% of principals move more than once, compared to 30% who move once. However, estimation of principal fixed effects does offer an advantage over estimation of teacher fixed effects in that principals are responsible for the performance of students in many grades. Thus, instead of looking at time-varying correlation in principal quality in order to estimate the true variation in principal effectiveness, I use cross-sectional variation. Specifically, I estimate Equation (5) separately for each grade to obtain an estimate $\hat{\mu}_{pg}$ of principal p's effectiveness in grade g. If grade-specific errors are independent so that $\hat{\mu}_{pg} = \mu_p^* + \nu_{pg}$, then $Cov(\hat{\mu}_{pg}, \hat{\mu}_{p,g-1}) = Var(\mu_p^*)$ where $\hat{\mu}_{p,g-1}$ is the estimate of principal p's effectiveness on the previous grade $g-1$. Thus, my estimate of the true variance of fixed effects is given by $\hat{\sigma}^2_{\mu^*} = Cov(\hat{\mu}_{pg}, \hat{\mu}_{pg-1})$. Thus, I construct

$$\hat{\lambda}_p = \frac{\hat{\sigma}^2_{\mu^*}}{\hat{\sigma}^2_{\mu^*} + \hat{\sigma}^2_\nu} \tag{9}$$

so that the adjusted fixed effect is given by:

$$VA_p = \hat{\lambda}_p \hat{\mu}_p \tag{10}$$

[21] For more discussion about the empirical content of value-added measures see Kane and Staiger (2008) and Rothstein (2009).

Estimating the variance of true principal ability across grades instead of across years credits principals for high performance that is common across grades. The downside of this approach is that it attributes school-wide common shocks not captured by the school fixed effect to principal performance. If, however, common shocks do indeed create bias in my shrinkage estimator, this bias should be greater in principal quality measured without school fixed effects at all. I check and find that it is not the case.

C Discretionary District Pay

The model presented in Section 3 assumes that schools do not have the flexibility to offer competitive wages to principals. Although the majority of principal pay is set at the state level, there is still the possibility that school districts may be compensating principals who work at poorly-performing schools after NCLB by altering discretionary salary supplements. Although I do not observe individual supplements, I have district-level data on expenditures for salary supplements and the percentage of principals receiving supplements. If principals are being compensated for increased probabilities of failure at certain schools, then supplements at school districts with more poorly-performing schools should rise relative to higher-performing districts in response to NCLB. This change can happen in two ways: first, average principal supplements can increase at poorly-performing districts; or second, if total district funds for supplements do not change, schools may want to reallocate supplements so that the percent of principals receiving supplements should differentially change. I do not find evidence for either of these district responses. In terms of both supplement size and distribution, districts with more schools likely to fail AYP do not seem to behave differently from low-failure districts. If anything, average supplement sizes tend to decrease at high-failure districts, suggesting that the change in the likelihood that a principal is subject to NCLB sanctions is not being fully compensated by pay changes. These results are reported in Appendix Table A.

D Specification Checks

A key concern in estimating Equation (7) is that VA_{pst} is only observed for principals who are movers in the pre-period, which introduces a potentially non-random missing data problem. To see this more clearly, suppose that principal p with observed value-added is observed at school s in year t, but moves to school s' in year $t+1$. In this case, school s is in my analytic sample in year t, but not in year $t+1$ because I am unable to observe the quality of the new principal at school s (unless the new principal it hires is one for which I have estimated value-added).

A concern for my empirical strategy is that, as a result of my sample construction, α_1 in Equation (7) may be capturing changes in the composition of schools I observe in my sample as opposed to true changes in the assignment of principals to schools arising from NCLB, because schools that retain their principal may be unobservably different from schools that do not. This fact alone, however, is not sufficient to generate bias in α_1.

For clarity, consider two schools, A and B, which are identical on observables and which initially both employ principals for whom I have observed value-added, but suppose that school B's principal leaves. In this case, I continue to observe school A's principal in the next year, but I no longer observe the quality of the next principal at school B. If school B's principal left for reasons related to the unobserved quality of the job at B, then the quality of the next principal at school B, which is unobserved, is likely to be different from the quality of the principal at school A, which is observed. This means that the average quality of principals who are observed in the sample is likely to be different from the true average quality of principals, for both the group of low- and high-risk schools. Yet, α_1 captures the difference-in-difference between quality at high- and low-risk schools, before and after NCLB. Thus, in order for this missing data issue to bias α_1, it must be that the bias in observed average quality 1) differs for high and low-risk schools and 2) changes after NCLB. If only 1) is true, then the bias introduced by the sample selection process is captured by the $\Pr(\text{fail})_s$ term in Equation (7). If only 2) holds, then these differences are captured by the time effects in Equation (7).

Conditions 1) and 2) both hold in one of two scenarios. Under the first, high- and low-risk schools must have different probabilities of leaving my sample and, in addition, there must be a change in the extent to which schools staying in the sample unobservably differ from schools that exit. If the degree of selection on unobservables changes, then the difference between observed average quality and true average quality would change before and after NCLB. If high and low-risk schools are equally likely to exit the sample, however, this bias is the same across $\Pr(\text{fail})_s$, so that it is captured by the time effects in Equation (7). Conversely, if the likelihood that high and low-risk schools leave the sample is different, then the bias in true and observed principal quality is likely to differ across these types of schools. If, however, there is no change in the degree of selection on unobservables, this bias does not change after NCLB and thus is captured by the $\Pr(\text{fail})_s$ term. A similar logic explains the second scenario leading to bias in α_1, which requires both that schools leaving the sample be unobservably different from those that stay, and that there be a differential change in the likelihood that high- and low-risk schools exit the sample after NCLB.

In the case where α_1 is biased, many sensible stories lead to α_1 being too small and bias me away from finding a negative allocative effect of NCLB on high-risk schools. For instance, suppose that after NCLB schools that lose their principals become more undesirable than observably identical schools that retain their principals. If undesirable schools have a harder time attracting high-quality principals, the average quality of principals in schools that are observed is likely to be higher than the true average quality, and this upward bias is likely to be larger the more missing observations there are. Thus if there is more turnover at high-risk schools (and thus more missing observations), the observed difference-in-difference in quality at low- and high-risk schools underestimates the true difference-in-difference because average quality at high-risk schools is more upwardly biased than average quality at low-risk schools.

In Appendix Table B, I examine the degree to which the probability of exiting my sample changes at low- and high-risk schools following NCLB and find no differential effect. There could still be bias if the selection on unobservables of principals as they leave schools changes after NCLB. I have no direct test for this, but in the remaining columns of Table B I show that there does not appear to be differential changes in the selection of schools out of the sample based on observables.

Another way of addressing this missing data problem is to examine only schools that employ a new principal with observed value-added in the next year in which the school is observed. With quality measurements of both the current and next principal, I can ask whether the next principal employed at high-risk schools is more likely to be lower quality after NCLB. In Appendix Table C, I examine the effect on NCLB on the next principal assigned to a school on a restricted sample where I observe both the quality of the current and future principal. This reduces the sample size by a significant amount, but I find evidence that both the math and reading quality of the next principal falls at high-risk schools following NCLB.

Appendix Table E reports results using alternative measures of school performance and principal quality. I find that I obtain similar results when measuring a school's exposure to NCLB by using the percentage of AYP targets it is likely to fail, and when I use unshrunken estimates of principal value-added. I find qualitatively similar but smaller and statistically insignificant effects of NCLB when principal value-added is measured using lagged test scores in the previous year. This result may reflect the fact that controlling for lagged test scores for the previous year does not fully credit a principal with cumulative test score gains made in her school.

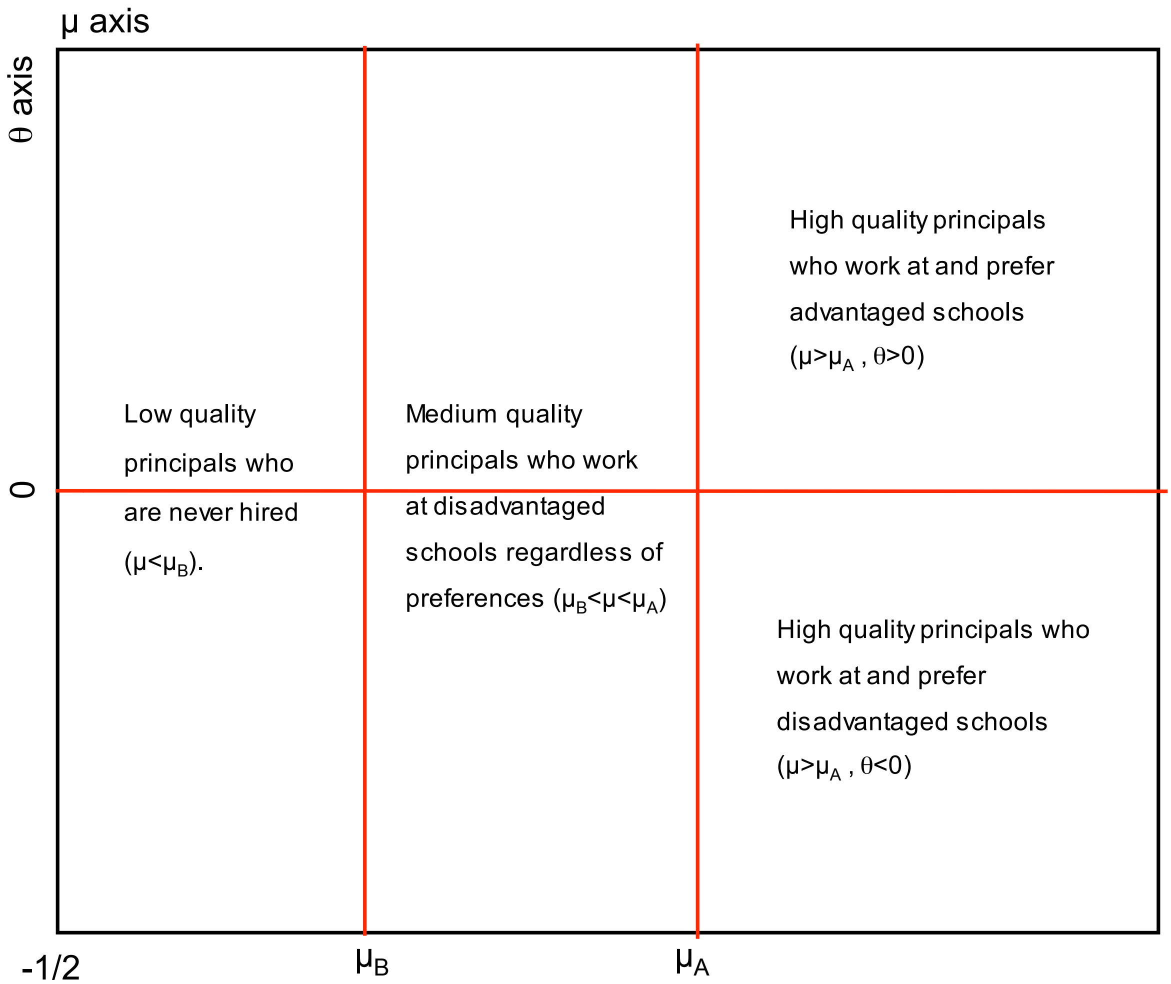


Figure 1: Initial allocation of principals

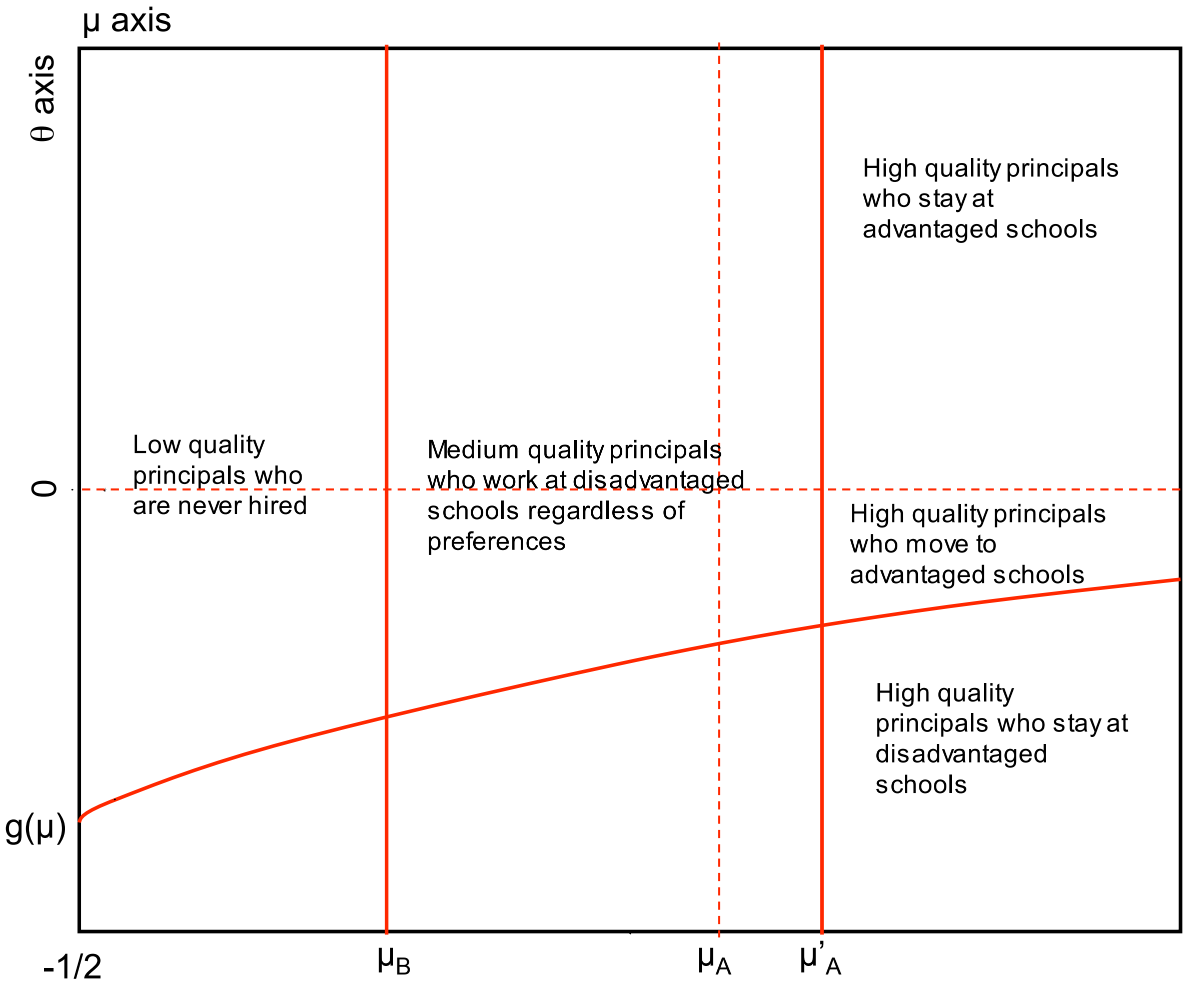


Figure 2: Allocation of principals after accountability

		# Schools	# Principals
	SAMPLE FOR ESTIMATING PRINCIPAL QUALITY: 1995-2002		
1	All full time principals in schools employing one principal at a time from 1995 to 2002.	2399	4890
2	Additionally: principals of schools including students with math and reading test scores.	2112	3030
3	Additionally: principals of schools that employ at least two principals from 1995 to 2002.	1200	2289
4	Additionally: principals of schools in grades 4-8, who have test scores for the current and previous year.	1097	2118
5	Additionally: principals of schools for which at least one principal is a mover between 1995 and 2002.	500	832
6	Additionally: principals for whom fixed effects are estimated.	500	640
7	Additionally: principals who are movers.	500	298
8	Additionally: principals for whom shrinkage can be computed.	500	275
	ANALYTIC SAMPLE: 1995-2007		
9	All schools at which mover principals with shrunken fixed effects work: 1995-2007.	596	298
10	Including only schools with standard grades, which are observed in all years.	383	214

Figure 3: Sample construction

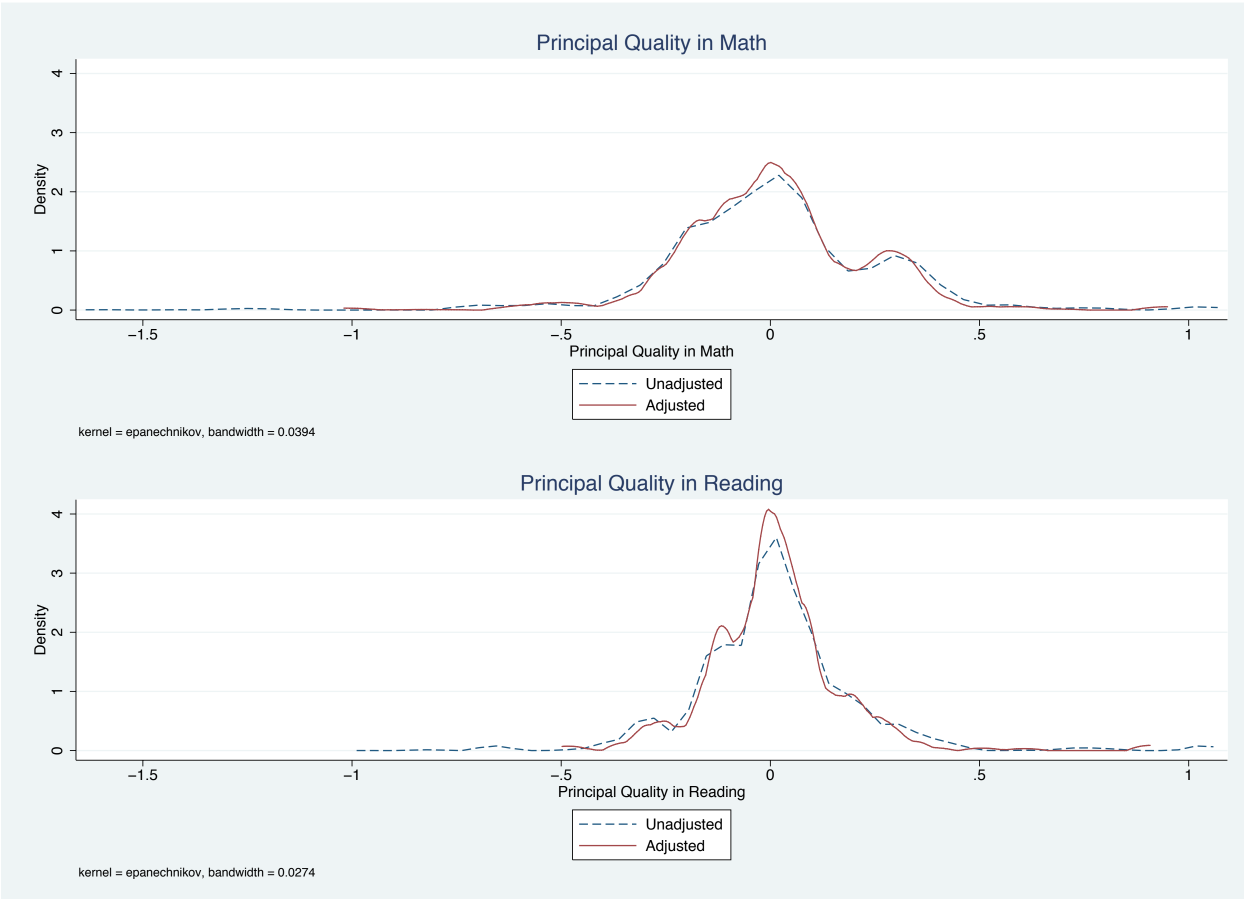
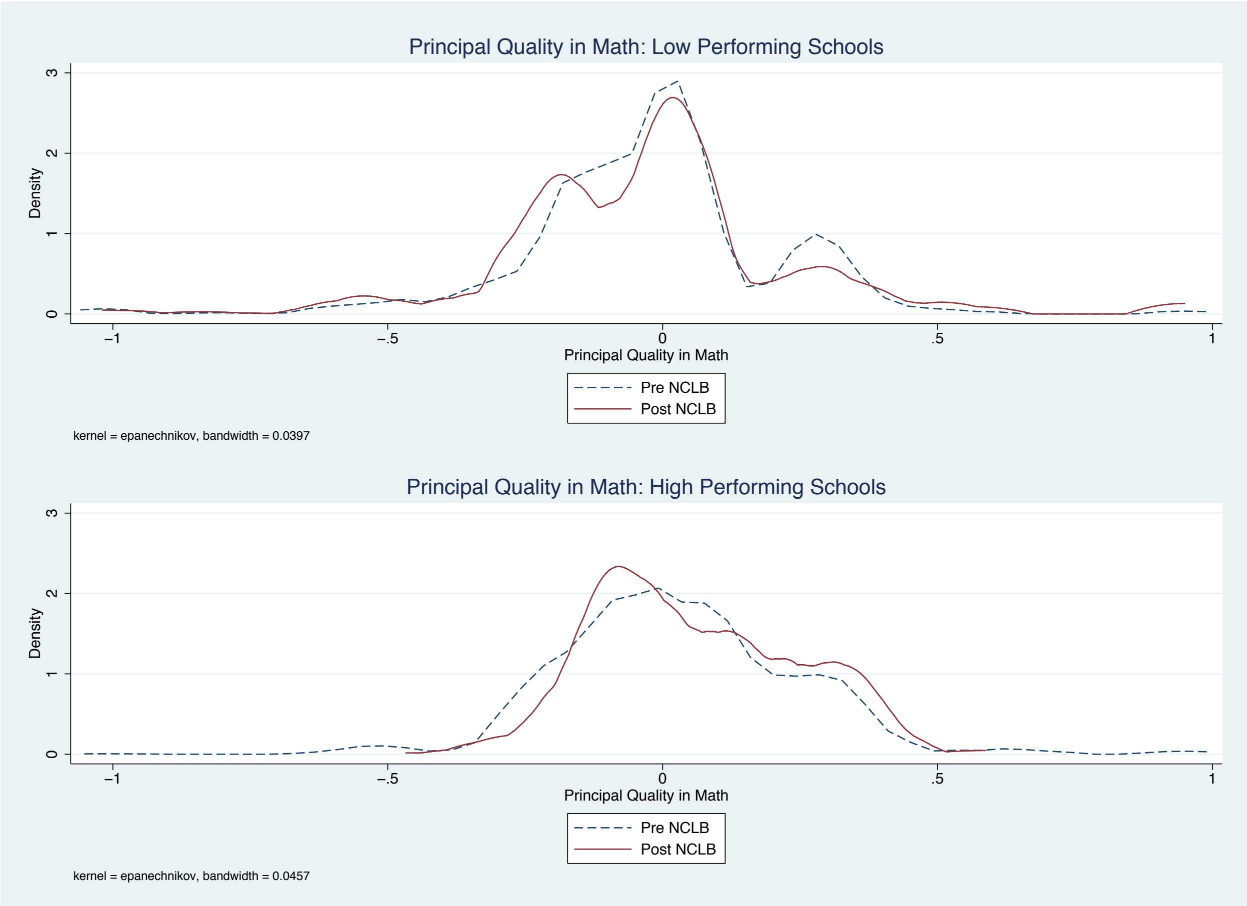


Figure 4: The Distribution of Principal Quality



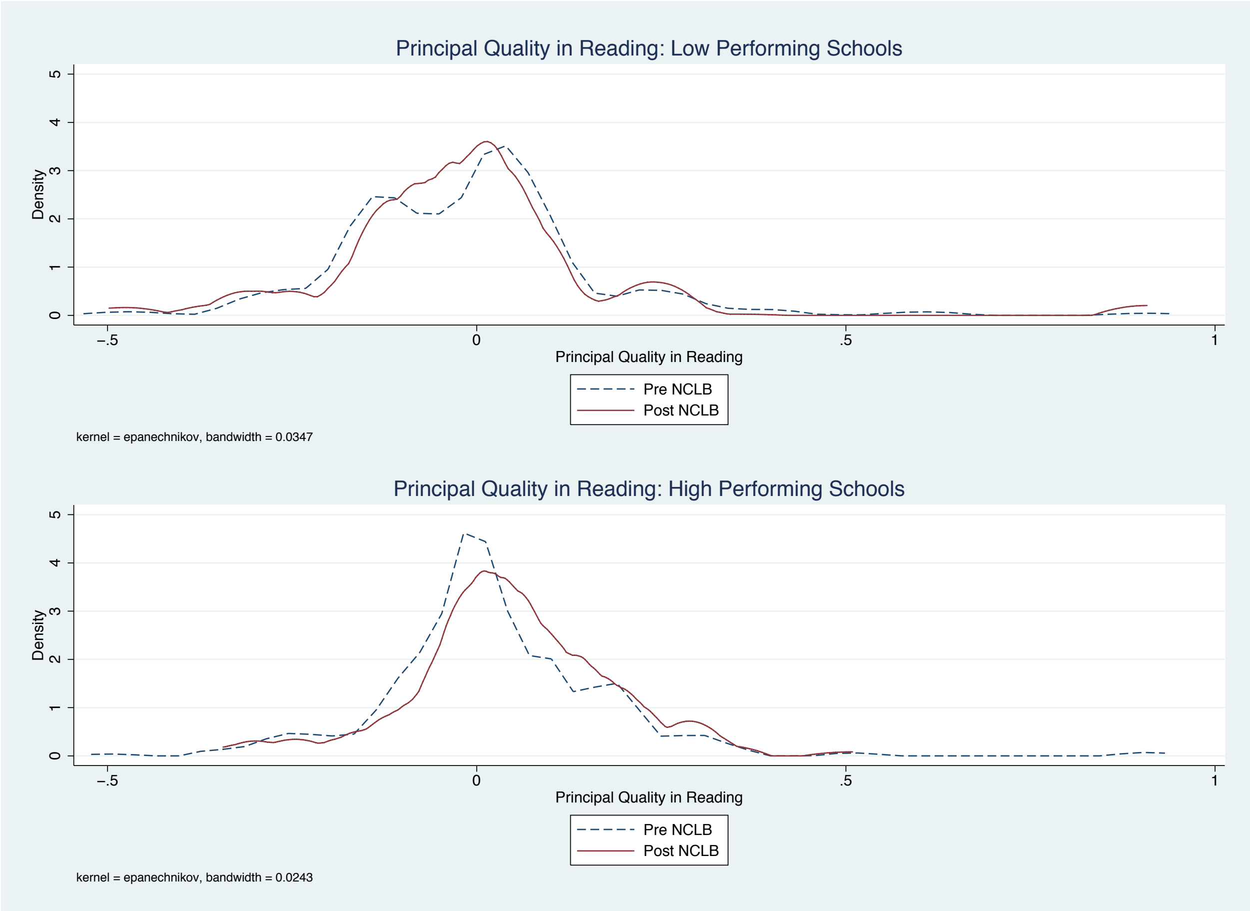
Percentiles of Principal Math Quality at Low-Performing Schools

	10	25	50	75	90	SD
Pre-NCLB	-0.236	-0.144	-0.006	0.054	0.284	0.233
Post-NCLB	-0.236	-0.17	-0.006	0.054	0.284	0.267

Percentiles of Principal Math Quality at High-Performing Schools

	10	25	50	75	90	SD
Pre-NCLB	-0.225	-0.104	0.023	0.142	0.308	0.211
Post-NCLB	-0.148	-0.073	0.05	0.179	0.322	0.184

Figure 5: The Distribution of Principal Math Quality Before and After NCLB



Percentiles of Principal Reading Quality at Low-Performing Schools

	10	25	50	75	90	SD
Pre-NCLB	-0.171	-0.12	0.003	0.062	0.118	0.159
Post-NCLB	-0.178	-0.098	-0.005	0.057	0.118	0.186

Percentiles of Principal Reading Quality at High-Performing Schools

	10	25	50	75	90	SD
Pre-NCLB	-0.116	-0.039	0.014	0.092	0.189	0.147
Post-NCLB	-0.116	-0.02	0.032	0.101	0.189	0.136

Figure 6: The Distribution of Principal Reading Quality Before and After NCLB

Figure 7: The Timing of NCLB's Effect on Principal Quality

TABLE 1: SUMMARY STATISTICS

	Sample	Universe	P-values
# Schools	383	1605	
# Principals	214	2054	
PRINCIPAL CHARACTERISTICS			
Years in data	10.1	8.4	0.000
	(2.29)	(2.84)	
% Ever switch schools	100	65.97	0.000
Years in data at first switch, conditional on switching	3.33	3.39	0.627
	(1.70)	(2.41)	
Imputed Age	48.31	46.81	0.119
	(6.45)	(8.08)	
Advanced Degree	36.85	34.17	0.914
State Salary	67,303	62,346	0.021
	(14,009)	(18,535)	
Principal Tenure (0/1)	0.74	0.77	0.000
Principal Experience (0/1)	0.90	0.87	0.298
SCHOOL CHARACTERISTICS			
Pr(Fail) AYP in 2002	56.15	50.47	0.002
	(41.69)	(42.68)	
% Urban	52.15	43	0.001
% Title 1	62.53	48.21	0.000
% Elementary School	72.03	48.7	0.000
Proportion Black	35.97	31.02	0.001
	(24.30)	(24.02)	
Proportion Hispanic	5.54	4.47	0.025
	(6.58)	(6.14)	
Proportion Asian	2.06	1.51	0.000
	(0.28)	(2.34)	
Proportion White	54.54	60.79	0.000
	(25.97)	(26.60)	
Proportion at Grade Level	76.09	75.29	0.359
	(11.84)	(12.17)	
Student-teacher Ratio	15.14	15.26	0.917
	(2.93)	(2.75)	
School Size	592	700	0.000
	(261)	(366)	
Joint			0.000

Notes: Standard devations are in parentheses. Observations are school-year cells. The universe sample include all schools that have 1) one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals who have worked at least one year before 2003. The analytic sample further restricts this sample to schools employing principals who 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects.

TABLE 2: PREDICTORS OF SAMPLE REPRESENTATION AND MOBILITY: ANALYTIC SAMPLE VS. UNIVERSE

	Total years observed	Years observed to date	Switch in next year	Retire in next year
	(1)	(2)	(3)	(4)
Panel 1: Are Mobility Patterns Different for Sample Principals on the Basis of Pr(Fail)?				
Pr(Fail) AYP in 2002 X 1(Sample Principal)	0.133	0.333	-0.027	-0.004
	(0.315)	(0.323)	(0.020)	(0.012)
Pr(Fail) AYP in 2002	-0.312*	-0.242**	0.030***	0.005
	(0.162)	(0.102)	(0.008)	(0.005)
1(Sample Principal)	1.347***	0.616***	0.076***	-0.023***
	(0.223)	(0.204)	(0.013)	(0.007)
Panel 2: Are Mobility Patterns Different for Sample Principals on the Basis of Student Race?				
Proportion White X 1(Sample Principal)	-0.361	-0.553	0.018	0.010
	(0.471)	(0.427)	(0.029)	(0.016)
Proportion White	0.678***	-0.423***	-0.048***	-0.068***
	(0.211)	(0.127)	(0.010)	(0.006)
1(Sample Principal)	1.668***	0.990***	0.048***	-0.032***
	(0.309)	(0.247)	(0.018)	(0.009)
Panel 3: Are Mobility Patterns Different for Sample Principals on the Basis of Student Income?				
Proportion Free/Reduced Lunch X 1(Sample	1.065**	0.159	0.045	-0.054**
	(0.489)	(0.550)	(0.037)	(0.025)
Proportion Free/Reduced Lunch	-1.287***	0.086	0.012	0.091***
	(0.245)	(0.165)	(0.014)	(0.010)
1(Sample Principal)	1.059***	0.641***	0.042***	-0.006
	(0.238)	(0.231)	(0.016)	(0.010)

Notes: Standard errors are in parenthesis. Each column in each panel is its own separate regression. Observations are school-year cells. The universe sample include all schools that have 1) one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals who have worked at least one year before 2003. The analytic sample further restricts this sample to schools employing principals who 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects. Total years refers to the total number of years a principal is observed in the NC state data, including years where she is employed by a school that is not always open or employs more than one principal. Years in observed to date, switching, and retirement are all defined on this extended sample.

TABLE 3: HOW IS PR(FAIL) CORRELATED WITH SCHOOL CHARACTERISTICS?

	Sample	Universe
Mean of Pr(Fail):	0.549	0.461
% Black	0.802***	0.722***
	(0.133)	(0.0637)
% Hispanic	1.048***	1.145***
	(0.322)	(0.140)
% Asian	1.606**	0.379
	(0.654)	(0.369)
% Other	0.666***	0.394***
	(0.199)	(0.142)
% Free Lunch	0.348*	0.504***
	(0.177)	(0.0905)
Students (1000s)	0.234**	0.396***
	(0.107)	(0.0501)
Elementary School	-0.141***	-0.158***
	(0.0486)	(0.0212)
N	1737	15337
R2	0.443	0.469

Notes: Pr(Fail) is the probability that a school fails AYP based on demographics, urbanicity, and school level from 1995 to 2002. This table does NOT report results from the actual estimation of Pr(Fail), which involves linear demographics for each year interacted with an elemenatary school dummy as well as yearly quadratics and cubics in student demographics. See text for details.

TABLE 4: CORRELATES OF PRINCIPAL QUALITY BEFORE AND AFTER NCLB

Dep. Var.	Math FE		Reading FE	
	Pre	Post	Pre	Post
	(1)	(2)	(1)	(2)
Pr(Fail)	-0.134***	-0.273***	-0.075**	-0.281***
	(0.049)	(0.060)	(0.030)	(0.062)
% At grade level	0.456***	0.577***	0.105	0.619***
	(0.156)	(0.200)	(0.107)	(0.204)
% White	0.077	0.264**	-0.030	0.289**
	(0.080)	(0.126)	(0.064)	(0.129)
% Black	-0.094	-0.306**	0.051	-0.338**
	(0.083)	(0.140)	(0.069)	(0.143)
% Hispanic	0.204	-0.117	-0.120	-0.122
	(0.317)	(0.421)	(0.196)	(0.430)
% Free Lunch	-0.059	-0.248**	0.004	-0.254**
	(0.090)	(0.110)	(0.062)	(0.115)

Notes: Each cell is a separate regression of the indicated variable on estimates of principal value-added, controlling for district fixed effects only, weighted by the inverse variance of the principal quality measure. Sample is the analytic sample of principals with estimated fixed effects. Standard errors are clustered at the school level.

TABLE 5: EFFECTS OF NCLB ON THE ALLOCATION OF PRINCIPAL QUALITY

	Principal Math Quality			Principal Reading Quality		
	All	Title I	Non Title I	All	Title I	Non Title I
	(1)	(2)	(3)	(4)	(3)	(4)
Pr(Fail) X 1(Year>2002)	-0.103**	-0.228***	0.039	-0.037	-0.104**	0.033
	(0.048)	(0.069)	(0.057)	(0.028)	(0.049)	(0.029)
Pr(Fail)	-0.152***	-0.113	-0.237***	-0.085***	-0.040	-0.108***
	(0.048)	(0.071)	(0.066)	(0.028)	(0.046)	(0.038)
Observations	1737	1120	617	1730	1117	613
R-squared	0.346	0.466	0.414	0.394	0.512	0.496

Notes: Sample is the set of schools that 1) employ one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals. Principals included must 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects. Pr(Fail) is the probability that a school will fail AYP based on prior demographics, urbanicity, and school level. 1(Year>2002) is a dummy for post 2002. Regressions control for district and year fixed effects, linear time trends for each district, and principal age and age squared. Standard errors are clustered at the school level. Regressions are weighted by the inverse variance of the relevant principal fixed effect.

TABLE 6: EFFECT OF NCLB ON THE ALLOCATION OF PRINCIPAL MATH QUALITY AT TITLE I SCHOOLS, BY DISTRICT CHARACTERISTICS

	Principal Math Quality			
	Large Districts	Small Districts	Urban	Non-Urban
	(1)	(2)	(3)	(4)
Pr(Fail) X 1(Year>2002)	-0.258**	-0.183**	-0.206**	-0.116
	(0.104)	(0.076)	(0.084)	(0.071)
Pr(Fail)	-0.168**	-0.045	-0.112	-0.228
	(0.080)	(0.100)	(0.072)	(0.148)
Observations	466	654	534	586
R-squared	0.409	0.612	0.462	0.597

Notes: Sample is the set of schools that 1) employ one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals. Principals included must 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects. Pr(Fail) is the probability that a school will fail AYP based on prior demographics, urbanicity, and school level. 1(Year>2002) is a dummy for post 2002. Regressions control for district and year fixed effects, linear time trends for each district, and principal age and age squared, with the exception of Column 3, which includes district fixed effects but not district-year linear time trends, due to the small sample size. Standard errors are bootstrapped and clustered at the school level. Regressions are weighted by the

TABLE 7: EFFECT OF NCLB ON AGGREGATE PRINCIPAL MOBILITY

	Sample - Switch	Universe - Switch	Universe - Retire
	(1)	(2)	(3)
Mean of dep. var.	0.153	0.096	0.089
Pr(Fail) X 1(Year>2002)	0.034	-0.011	0.004
	(0.050)	(0.014)	(0.016)
Pr(Fail)	-0.032	0.028***	0.006
	(0.027)	(0.009)	(0.006)
Observations	1670	11280	11280
R-squared	0.111	0.038	0.113

Notes: The universe sample include all schools that have 1) one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals who work at least one year prior to 2003. The analytic sample further restricts this sample to schools employing principals who 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects. This sample is smaller than the main sample because I exclude the year 2007 in order to observe where principals end up at in the next year. Switch is an indicator for whether a principal becomes a principal at a different school in the following year. Retire is a dummy equal to one if the principal is no longer working as a principal in the following year. Pr(Fail) is the probability that a school will fail AYP based on prior demographics, urbanicity, and school level. 1(Year>2002) is a dummy for post 2002. Regressions control for district and year fixed effects, linear time trends for each district, and principal age and age squared. Standard errors are clustered at the school level.

TABLE 8: EFFECT OF NCLB ON PRINCIPAL-SCHOOL MATCHING, BY PRINCIPAL MATH QUALITY

		Characteristics of the school to which a principal moves, conditional on moving			
	Switch	Pr(Fail)	% At Grade Level	% White	Title I
	(1)	(2)	(3)	(4)	(5)
Mean of Dep. Var.	0.153	0.555	0.755	0.536	0.537
Pr(Fail) X 1(Year>2002) for Math Quality < Median	0.085	-0.341	-0.071	-0.039	-0.179
	(0.076)	(0.362)	(0.095)	(0.185)	(0.395)
Pr(Fail) X 1(Year>2002) for Math Quality > Median	-0.019	-0.723	0.219**	0.555***	-0.851**
	(0.066)	(0.447)	(0.106)	(0.194)	(0.428)
Observations	1670	213	235	254	252
R-squared	0.112	0.701	0.681	0.665	0.577

Notes: Reported are coefficients on Pr(fail)X1(Year>2002) for each ability group. Sample is the set of schools that 1) employ one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals. Principals included must 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects. Pr(Fail) probability of failing AYP based on prior demographics, urbanicity, and school level. 1(Year>2002) is a dummy for post 2002. 1(Year>2002) is a dummy for post 2002. Regressions control for district and year fixed effects, linear time trends for each district, and principal age and age squared. All regressions include year and district fixed effects and district by year linear trends. Standard errors are clustered at the school level.

TABLE 9: EFFECT OF NCLB ON PRINCIPAL-SCHOOL ASSIGNMENT: SCHOOLS EXPOSED TO "SHOCK" RELATIVE TO ABC PERFORMANCE

PANEL 1: DISTRIBUTION OF QUALITY

	Principal Math Quality		
	All	Title I	Non Title I
Dep. Var.	(1)	(2)	(3)
Shocked X 1(Year>2002)	-0.175**	-0.329**	0.019
	(0.074)	(0.129)	(0.070)
Shocked	-0.027	0.087	-0.171*
	(0.046)	(0.066)	(0.097)
Observations	1791	1120	671
R-squared	0.313	0.456	0.362

	Pr(Fail)	% At Grade Level	% White	# School-Years
Shock=1	94.88	69.24	33.7	224
Shock=0	50.41	77.06	57.52	1,567

PANEL 2: MOBILITY

		Characteristics of the school a principal moves to, conditional on moving		
	Switch (0/1)	% At Grade Level	% White	Title I
	(4)	(5)	(6)	(7)
Shocked X 1(Year>2002) Math Quality < Median	0.028	-0.047	-0.123	-0.398
	(0.074)	(0.077)	(0.202)	(0.412)
Shocked X 1(Year>2002) Math Quality > Median	0.015	0.127**	0.242	-0.735
	(0.086)	(0.060)	(0.174)	(0.460)
Observations	1714	241	263	261
R-squared	0.118	0.668	0.644	0.585

Notes: A "shocked" school is one which was rated exempary or expected under the ABCs in 2002, but which failed over 25% of AYP categories in 2002-03. The sample is the set of schools that 1) employ one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals. Principals included must 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects. 1(Year>2002) is a dummy for post 2002. Regressions control for district and year fixed effects, linear time trends for each district, and principal age and age squared. Standard errors are clustered at the school level. Regressions are weighted by the inverse variance of the relevant principal fixed effect.

APPENDIX TABLE A: EFFECT OF NCLB ON DISTRICT SALARY SUPPLEMENTS

	Sample		Universe	
	% Receiving	Avg. supplement (1000s)	% Receiving	Avg. supplement (1000s)
	(1)	(2)	(3)	(4)
Mean of dep. var.	0.968	7.031	0.929	5.841
Pr(Fail) X 1(Year>2002)	-0.012	-2.063	-0.001	-2.288
	(0.038)	(2.228)	(0.047)	(1.686)
Pr(Fail)	-0.012	8.557*	0.065	7.486*
	(0.048)	(4.961)	(0.070)	(4.042)
Observations	320	320	667	667
R-squared	0.011	0.078	0.014	0.046

Notes: Regression is at the district-year level for the years 2002-2007. % Receiving indicates the percentage of principals receiving a district supplement, average supplement includes zeros, in 2007 dollars. Year fixed effects are included. Standard errors are clustered at the district level and the regression is weighted by district size.

APPENDIX TABLE B: DIFFERENTIAL CHANGES IN SAMPLE EXIT BY SCHOOL CHARACTERISTICS

	RHS School Characteristics			
Dep. Var.: 1(School leaves sample in the next year)	Pr(Fail)	% Grade Level	% White	% Free Lunch
	(1)	(2)	(3)	(4)
FULL SAMPLE				
School Characteristic X 1(Year>2002)	0.043	-0.216	0.112	0.047
	-0.066	(0.306)	(0.106)	(0.107)
School Characteristic	-0.025	-0.121	-0.093	0.008
	-0.031	(0.135)	(0.065)	(0.062)
Observations	1670	1502	1714	1707
R-squared	0.117	0.127	0.125	0.123
TITLE I ONLY				
School Characteristic X 1(Year>2002)	0.055	-0.068	0.164	-0.002
	(0.091)	(0.364)	(0.160)	(0.151)
School Characteristic	0.098	-0.231	-0.250**	0.110
	(0.061)	(0.201)	(0.102)	(0.094)
Observations	1075	939	1075	1068
R-squared	0.168	0.168	0.171	0.167

Notes: Sample is the set of schools that 1) employ one principal at a time; 2) are open in all years between 1995 and 2007; and 3) employ full-time principals. Principals included must 1) work at least two elementary or middle schools before 2003, and 2) have estimated fixed effects. Pr(Fail) is the probability that a school will fail AYP based on prior demographics, urbanicity, and school level. 1(Year>2002) is a dummy for post 2002. Regressions include year fixed and district fixed effects, principal age and age squared, and district by year linear trends.

APPENDIX TABLE C: EFFECT OF NCLB ON THE QUALITY OF THE NEXT PRINCIPAL WORKING AT A SCHOOL, RESTRICTED SAMPLE

	Math Quality of the Next Principal	Reading Quality of the Next Principal
	(1)	(2)
Pr(Fail) X 1(Year>2002)	-0.864	-0.426
	(0.968)	(2.598)
Pr(Fail)	0.398	1.689
	(0.715)	(1.091)
Observations	56	34
R-squared	0.789	0.949

Notes: Sample includes only school-year observations for which a school changes principals in the next year in which it is observed, and for which I observe the new principal's estimated fixed effect. The dependant variable is the quality of the next principal.

APPENDIX TABLE D: EFFECT OF NCLB ON THE ALLOCATION OF PRINCIPAL MATH QUALITY WITH ALTERNATIVE SPECIFICATIONS

	Alternative Principal Quality Measure: Using lagged test scores to measure FE		Alternative Failure Probability Measure: using predicted % of targets failed		Alternative Principal Quality Measure: using unshrunken fixed effects	
	Title I	Non Title I	Title I	Non Title I	Title I	Non Title I
Math FE	(1)	(2)	(3)	(4)	(5)	(6)
Pr(Fail) X 1(Year>2002)	-0.076	0.006	-0.276**	-0.049	-0.251***	0.045
	(0.049)	(0.046)	(0.138)	(0.080)	(0.075)	(0.064)
Pr(Fail)	-0.063	-0.134**	0.041	-0.109	-0.122	-0.259***
	(0.049)	(0.067)	(0.091)	(0.095)	(0.077)	(0.075)
Observations	1109	624	1120	671	1120	617
R-squared	0.509	0.365	0.442	0.348	0.469	0.372

Notes: Sample includes observations for all schools employing principals who were movers prior to 2003, and for whom I have estimated fixed effects. Sample splits are based on Title I status in 2002-03. Regressions in the first panel use math fixed effects computed with lagged student test scores. School performance is defined as the probability that a school will fail AYP based on pre-period demographics. Regressions in the second panel are based on school performance measured as percent of AYP targets a school is predicted to fail based on pre-period characteristics. Regressions in the final panel use unadjusted fixed effects. All regressions control for district and year fixed effects, linear time trends for each district, and principal age and age squared. Standard errors are clustered at the school-year level. Regressions are weighted by the inverse variance of the relevant principal fixed effect. Standard errors are clustered at the school level.