Testing the No Child Left Behind Act: analyzing assumptions about the link between teacher quality and student achievement

Trinh Tran
Department of Sociology
University of California, Berkeley
ttran1@berkeley.edu

September 2008

Abstract

Teachers have long been considered a critical educational resource. The No Child Left Behind Act (NCLB) of 2001 recognizes the importance of educators through its highly qualified teachers mandate. Under NCLB, teachers are deemed highly qualified if they (1) have a bachelor's degree, (2) possess full state teacher certification, and (3) demonstrate knowledge in the subjects taught. NCLB assumes that teachers who meet highly qualified status will also produce the greatest gains in student learning. Using panel data analysis and the Los Angeles Unified School District, this study asks (1) how has NCLB changed the composition of teachers' qualifications? and (2) what effect do NCLB-defined teacher qualifications have on student achievement? The data show that NCLB has succeeded in changing the composition of teachers' qualifications. Beginning with the law's enactment, schools altered their teacher workforce to more closely reflect NCLB standards of teacher quality. Contra NCLB's assumptions, the data show a negative relationship between increasing the proportion of teachers with additional years of schooling and student growth. The data show no statistically meaningful relationship between teachers who are fully certified and student achievement.

Introduction

Teachers have long been considered a critical educational resource. Indeed, schools allocate the bulk of their budget to teachers' salaries (Hanushek and Rivkin 2003). Thus, it makes sense that the No Child Left Behind Act (NCLB) of 2001 would recognize the importance of educators through its highly qualified teachers provision. At the heart of NCLB's teacher mandate is the theory that raising the quality of the teacher workforce will subsequently raise student achievement. The legislation establishes the first set of federal guidelines ever for teacher quality. NCLB considers teachers highly qualified if they (1) have a bachelor's degree, (2) possess full state teacher certification, and (3) demonstrate knowledge in the subjects taught.[1]

NCLB's approach to raising the quality of the teacher workforce dramatically departs from earlier federal initiatives. Whereas prior efforts included grants to improve the training of prospective teachers and demanded accountability from collegiate schools of education, NCLB focuses on hiring qualified teachers and places this responsibility on schools and districts (United States Government Accountability Office 2007). Yet NCLB's definition of highly qualified teachers masks the wide disagreement among scholars over the traits and qualifications of a highly *effective* teacher. NCLB assumes that teachers who meet highly qualified teacher status will also produce the greatest gains in student learning. However, studies that explore the relationship between teacher traits and student achievement question the very criteria NCLB uses to determine teacher quality. These studies debate whether and which paper-based measures of teacher quality can adequately predict student performance (Goldhaber and Brewer 2000; Kane *et al.* 2005; Gordon *et al.* 2006).

[1] Appendix Table A1 discusses in more detail California's NCLB teacher requirements.

In this paper, I investigate NCLB's assumptions regarding the relationship between teacher quality and student achievement. I ask two related questions: (1) How has NCLB affected the composition of teachers' qualifications? and (2) What effect do NCLB-defined teacher qualifications have on student achievement?

These questions matter on both substantive and theoretical grounds. Substantively, this study investigates the consequences of policy and asks whether paper-based qualifications can adequately identify effective teachers. NCLB presupposes that we can determine teacher effectiveness at the point of hiring. It assumes that we can screen teachers based on qualifications like whether they have a college degree, are fully credentialed, and possess subject matter expertise. NCLB shifts the focus away from other initiatives that have attempted to improve teacher quality through pre-service and ongoing professional development. The legislation instead assumes that teachers gain all the skills and knowledge they need to be effective on-the-job before the first day of school.

Theoretically, this study seeks to understand whether the high-stakes accountability reform movement, of which NCLB is today the main driver, contributes to educational achievement. NCLB uses a carrot-and-stick approach to encourage growth in student performance. The initiative assumes that schools, teachers, and students will not perform as well as they could without the threat of sanctions and the promise of rewards. The legislation presupposes that schools are not already performing at capacity. Additionally, the law assumes that increased external oversight will help schools achieve those goals. NCLB's definition of a highly qualified teacher ignores the fact that schools may have better internal ways of evaluating teachers (e.g., through principal, peer, and

parent evaluations). NCLB's teacher mandate assumes that schools need an external set of guidelines to help them weed out ineffective teachers.

In the following sections, I first describe the historical context and theoretical basis upon which NCLB was founded. Second, I discuss existing research that explores the relationship between teacher quality and student achievement. Finally, I present results from my analysis of the relationship between NCLB, teacher quality, and student achievement within the Los Angeles Unified School District.

The No Child Left Behind Act (2001)

Historical context and theoretical background

To understand the relationship between NCLB, teacher quality, and student achievement, we must first examine the historical context and theoretical basis upon which the legislation was founded. The theory behind NCLB—that rewards and sanctions can motivate educational excellence—is deeply rooted in the historical practice of using tests to evaluate not just students but also schools and teachers.

Testing programs have long figured centrally in American public education. Ravitch (2002) traces their use to nineteenth century school and college administrators who promoted and admitted students based solely on test results. However, Ravitch states that the idea of linking testing with accountability—the notion of holding students, teachers, principals, schools, and districts liable for student performance—is "*not* [my emphasis] one with a long pedigree" (9). For the most part, testing remained unconnected to school and teacher accountability during the nineteenth century. Teachers of this period were only answerable to certification tests (Ravitch 2002). These tests aimed to measure subject knowledge and usually included an interview on religious

beliefs. Once admitted into the profession, teachers were no longer questioned about their competence or suitability. Students were blamed for their own failure. They alone suffered the consequences, either by failing or dropping out.

Beginning in the early twentieth century with the establishment of the field of educational psychology, testing began to form a closer relationship to school and teacher performance. Edward L. Thorndike, a preeminent educational psychologist at Teachers College, Columbia University, led the push to make education an exact science. In his effort to apply scientific rigor to education, he developed standard scales for testing pedagogical methods and subject knowledge. Thorndike did not design testing for the purposes of external control. Rather, he believed that through measurement-driven instruction, educators could improve professional practice, thereby retaining professional oversight of schools and limiting state and outside intrusion (Ravitch 2002).

Despite his intentions, Thorndike laid the groundwork for the current accountability movement. Standardized testing as a mechanism for gauging and improving performance became ever more popular with the establishment of Title I of the Elementary and Secondary Education Act of 1965 (Heubert and Hauser 1999).[2] From then on, the federal government increasingly relied upon large-scale testing to evaluate programs supporting the education of children in underserved communities. Testing took on even greater importance during the minimum competency movement of the 1970s. This time, the idea of attaching high stakes (i.e., rewards and sanctions) to test scores gained momentum as students (and in some cases schools) were penalized for their failure to master basic skills. Heubert and Hauser (1999) explain the two fundamentally different ways tests have been used to further policy goals: "A low-stakes test has no

[2] Title I earmarks federal funds for economically disadvantaged students.

significant, tangible, or direct consequences attached to the results, with information alone assumed to be a sufficient incentive for people to act…In contrast, high-stakes policies assume that information alone is insufficient to motivate educators to teach well and students to perform to high standards. Hence, it is assumed, the promise of rewards or the threat of sanctions is needed to ensure change" (35).

The reliance on testing as a high-stakes accountability mechanism continued to grow with the 1983 publication of *A Nation at Risk* (National Commission on Excellence in Education), which forecast economic doom because of poor education (Pipho 1985). During this time, many state governors looked towards business to help reform education. Ravitch (2002) reports that business leaders encouraged governors and other elected officials to adopt business incentive structures to enhance academic motivation. Additionally, the standards-based movement, which emerged in the 1990s, supported the alignment of testing, standards, and curriculum, combined with rewards and sanctions, to motivate student learning (Smith *et al.* 1990).

Thus, while testing has a long history in the American public school system, its use as a mechanism for holding schools and teachers accountable is fairly new. Even more recent is the attachment of high stakes to test scores. The high-stakes accountability reform movement continues to exert a strong presence today in the form of NCLB, which mandates annual testing in grades 3-8 and 10-12 in return for Title I funding.[3] Indeed, the law faithfully adheres to the high-stakes accountability movement by integrating standards, testing, and rewards/sanctions.

Yet NCLB goes one step beyond the high-stakes accountability movement. It seeks to codify and quantify not only student learning but also the standards and

[3] Appendix A discusses the legislation in more detail.

measures of teacher quality. As mentioned earlier, previous federal education laws sought to improve teacher quality by reforming teacher preparation programs and offering better training and support for teacher candidates. NCLB, however, focuses on hiring a new crop of teachers. The legislation establishes the first set of federal guidelines through which teachers are deemed highly qualified. As the next section makes clear, scholars who work under the education production tradition hotly contest NCLB's teacher guidelines.

Education Production Models

Ever since Coleman *et al.'s* seminal study *Equality of Educational Opportunity* (1966), social scientists and economists have been interested in the link between school inputs and student outputs. Coleman *et al.* (1966) maintained that "schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (325). The report concluded that financial inputs to a school system had no effect on its overall educational output. This controversial finding, which has been alternately refuted and confirmed, has motivated a wealth of scholarship that focuses on the link between school resources and student achievement.

The education production tradition is one example of research inspired by the Coleman report. This approach uses the metaphor of the factory and applies economic concepts to explain the link between school resources and student outcomes; it views schools as producers of student achievement (Greenwald *et al.* 1996). Education production scholars generally focus on (1) the specific school-level resources that promote student achievement and (2) the magnitude of school-level resources on student achievement. No single education production model exists as scholars disagree over

which school inputs matter and how to operationalize this concept. However, common school-level resources considered include per student expenditure, teacher characteristics, and class and school size. Student outputs usually include standardized test scores.

Even though the education production literature is large and active, the field is troubled. Scholars disagree over how to identify and measure school-level variables (i.e., the most appropriate methodology) and the magnitude of school-level effects. Studies on the effects of teacher characteristics, an important subset of the education production literature, illustrate the varied methodological approaches researchers have used and the inconsistent findings that plague the field.

Within the teacher effects literature, the methodological debate centers primarily on the appropriate level of analysis. Researchers have examined the relationship between teacher inputs and student outputs at the state, district, school, and classroom levels. Each level poses analytic problems. State-level analyses, for example, are subject to aggregation bias. These models fail to capture variations in state policies that may influence the distribution of teacher characteristics (e.g., teacher certification requirements, collective bargaining rules on teacher contracts, tax base structures for public school funding). In contrast, single state analyses eliminate the largest sources of bias as the policy environment is held constant. District- and school-level studies also produce biased estimates as they do not take into account the entire stream of inputs (e.g., school-level resources, family, peer, and community influence) that happen throughout an individual student's school career. Value-added approaches, which examine gains in test scores of individual students from one grade to the next, erroneously assume that students and teachers are randomly assigned to one another. They fail to consider that

teachers are frequently assigned to students based on student characteristics (e.g., achievement). A more experienced teacher may be assigned to a class of high-achieving students because of seniority privileges. Conversely, schools may assign underperforming students to more experienced teachers as a compensatory strategy. The causal direction between teacher experience and student achievement in both examples is reversed.

While they differ over methodological strategies, education production scholars generally concur on which teacher variables should be indexed. They analyze those teacher traits that can be both measured and manipulated in one way or another by policy. Education production analysts rarely examine hard-to-measure and hard-to-control teacher quality variables such as career dedication and classroom control. Instead, they are largely concerned with traits—such as teacher education and years of experience—that influence policy decisions (e.g., teacher hiring, pay, and seniority). Thus, the six teacher resources most commonly analyzed—teacher verbal ability, teacher academic ability, teacher experience, teacher subject matter expertise, teacher certification, and teacher degree—share the attributes of being both measurable and policy relevant. Yet, as the next section details, while education production scholars agree on which teacher variables to analyze, they vehemently disagree over which traits matter most and by how much.

Teacher verbal ability

Of the teacher inputs examined, the most consistent findings emerge from research on teacher verbal ability. These studies use teacher performance on verbal ability tests as proxies for teachers' communicative skills. From his study of almost 900

districts in Texas, Ferguson (1991) concluded that a teacher's language skill as measured by performance on a statewide recertification exam is the most significant teacher characteristic for raising both math and reading scores. He found that teachers' verbal ability scores explained between twenty and twenty-five percent of the variation across districts in students' test scores. Similarly, Hanushek (1971) found that teachers' scores on a verbal facility test correlate positively with student achievement. In a later study, however, Hanushek (1992) found that teachers' performance on word tests affected their student's reading score gains but not their vocabulary score gains. Hanushek and Rivkin (2003) caution that because the focus and content of both teacher and student tests differ from one to the next, researchers may be capturing the effects of some other unknown teacher trait.

Teacher academic ability

Researchers have also studied the relationship between teachers' academic abilities and student test gains. These studies generally find that teachers who rate high academically (as measured by attendance at selective undergraduate institutions) are more effective in the classroom. In their analysis of 627 sixth-grade Philadelphia elementary school students, Summers and Wolfe (1977) used college quality as a measure of teachers' academic ability. They used the Gourman rating of each teacher's undergraduate institution to determine college quality.[4] They found that sixth-graders' scores on the Iowa Test of Basic Skills were significantly and positively related to the quality of a teacher's undergraduate institution. Additionally, they found that students from lower income families benefited the most from teachers who receive degrees from

[4] Gourman's ranking system relies on data from the institution's facilities, departments, administration, faculty, and alumni.

higher-rated colleges. Ehrenberg and Brewer (1981) also found that teaching effectiveness correlates positively with selectivity of teacher's undergraduate institutions. Using data from the High School and Beyond survey, a longitudinal survey of tenth and twelfth graders, they discovered that both white and black high school students benefited from teachers who graduated from better-rated undergraduate institutions as measured by Barron.[5] The results for Hispanic students were inconclusive.

Teacher experience

The research on teacher experience generally show threshold, non-linear effects. Teachers perform better (i.e., raise student scores) with a few years of experience but after a while this effectiveness drops. Murnane and Phillips (1981) found that teachers become more skilled after the first year of teaching. Rivkin *et al.* (2005) analyzed separately the effects of teacher experience for teachers who leave the profession vs. those who remain. They discovered that teachers who learn to teach better in the first three years of their career improve students' scores. However, after this initial period, student scores show little additional improvement.

Teacher subject matter expertise

Studies using different proxies for teacher subject matter knowledge have produced mixed results. Summers and Wolfe (1977) found a negative relationship between teachers' scores on the subject matter tests of the National Teacher Examinations and sixth-grade student test scores. In contrast, Goldhaber and Brewer (2000), using data from the National Educational Longitudinal Study of 1988, found that eighth-grade math students who have teachers with a bachelor's or master's degree in

[5] Barron ranks colleges using data from the entering class' entrance test scores and high school records. The rating system also takes into account the percentage of admitted applicants.

math have higher test scores relative to those whose teachers have out-of-subject degrees. This finding is consistent with Monk's (1994) analysis of the Longitudinal Study of American Youth, a survey of middle school and high school math and science education. He found that teacher content preparation, as measured by the number of courses a teacher took in the subject being taught, is positively related to how much science or math high school students learn. This relationship is curvilinear but still monotonic—returns from teacher's coursework on student achievement diminish after the teacher has taken four to six courses in the subject.

Teacher certification

Education production scholars also debate the extent to which traditionally certified teachers raise student achievement. Using state-level data, Darling-Hammond (1999) found that the percentage of teachers with full certification and college majors in the subject correlates positively with student achievement in reading and math, both before and after controlling for student poverty and language status. Additionally, she discovered that states with high growth in student achievement have enacted and maintained higher standards for entry to teaching while many states with low growth have not.

In a separate analysis, Darling-Hammond *et al.* (2005) found that students assigned to uncertified teachers performed .5 percentile points worse on an achievement test than those assigned to traditionally certified teachers. Students assigned to alternatively certified teachers (i.e., emergency, provisional, or temporary certification) underperformed by 2.5 points. In a state-level analysis of California schools, Betts *et al.* (2000) similarly found significant, negative relationships between average student scores

on state tests and the percentage of teachers on emergency permits. Likewise, in their study of New York City public school teachers, Kane *et al.* (2005) found that the average traditionally certified teacher raised reading scores one percentile point more than the average alternatively certified teacher. However, they point out that this one percentile point difference between certified and uncertified teachers is dwarfed by the differences within groups. That is, the substantive significance of teacher certification is low even if statistical significance is high. They conclude that traditional certification does not guarantee teacher quality. Goldhaber and Brewer (2000) share Kane *et al.'s* reservations regarding conventional routes to teacher certification. They found that math and science students who have teachers with emergency credentials do no worse than students whose teachers hold standard teaching credentials.

Teacher degree

Scholars also reach different conclusions on the relationship between teacher's level of education and student outcomes. Research that find a positive, significant effect also find that teacher's educational degree matters less than other measures of teacher quality. Ferguson (1991) found that teachers with a master's degree have significant, positive effects on student test scores but that this trait matters less than teachers' language skill and experience. He found that the percentage of teachers who have master's degrees accounts for about five percent of the variation in student scores across districts for grades one through seven. Master's degrees lose predictive power after the seventh grade. Darling-Hammond (1999) found that teachers' education level, while positively correlated, is a less powerful predictor of student achievement than whether a teacher is certified and holds a degree in the subject to be taught.

Teacher-effects literature conclusion

Even while the literature shows promising links between teacher verbal and academic ability and student achievement, we should not overstate the magnitude of these effects. As Hanushek and Rivkin (2003) argue, the focus and content of teacher verbal and student tests vary greatly from one study to the next. We need to first align what these assessments cover before drawing strong conclusions. Similarly, the findings on teacher's academic abilities encourage further research that use multiple measures of college rigor. The Gourman and Barron ranking systems differ markedly; it would be useful to know if they generate the same conclusions regarding teacher academic ability and student performance.

Findings regarding teacher experience also deserve further consideration as teacher experience may be confounded with other attributes that have nothing to do with that particular trait. For example, it is difficult to parse out cohort effects (i.e., the possibility that teachers who enter the profession at the same time tend to share certain common traits) from teacher experience. Cohorts of teachers hired in times of shortage may be less qualified than those hired when schools can be more selective (Darling-Hammond 1999). Measuring the effects of teacher experience is also difficult because many senior teachers can choose to teach in better schools.

We can offer several explanations for the varied and sometimes contradictory findings regarding teacher subject matter expertise, teacher certification, teacher degree, and teacher experience. One possible reason for contradictory findings regarding subject matter competency may have to do with the different ways analysts operationalize this variable. Summers and Wolfe (1977) used teachers' scores on subject tests as proxies for

subject matter knowledge. This may be a crude measure of subject matter expertise. A teacher's degree and coursework in a particular subject—the proxies used by Goldhaber and Brewer (2000) and Monk (1994) respectively—may more accurately reflect the extent of knowledge a teacher has in that particular subject than a one-time subject matter test.

Contradictory findings regarding full certification may be due to the great variation in licensing and certification standards among states. State requirements for admission into teacher education programs and/or teacher credentials differ along several dimensions, including degrees, coursework, and test scores.[6] Thus, it may not be enough to distinguish between the types of certifications teachers hold (i.e., full, emergency, provisional, or temporary). Rather, it may also be necessary to compare the rigor of certification. Those states that impose high cut-off scores on teacher exams and minimum GPA, demand more intensive coursework in the subject(s) to be taught, and include extensive student teaching training may generate higher student gains.

Finally, the different findings on the value of a master's degree may be due to the lack of distinction made in the type of master's degree held. Teachers who hold a master's degree in the subject to be taught may be more effective than teachers who hold an out-of-field degree. Similarly, a master's degree may not matter unless it is a master's in education science.

[6] Most states specify a minimum grade point average and/or minimum scores on tests of basic skills, general academic ability, or general knowledge for admittance to teacher education programs or to earn a teaching credential. Many states also require that potential teachers complete a major or minor in the subject(s) to be taught and/or pass a subject matter tests, take education classes, and/or pass a test of teaching knowledge and skill. While there is some commonality among state certification standards, there is also great variation. In 2001-2002, 37 states required tests of basic skills or general knowledge, 33 states required tests of content knowledge, and 26 required tests of pedagogy (Wayne and Youngs 2003).

Hypotheses

As the most recent heir to both the high-stakes accountability and standards-based reform movements, NCLB makes several assumptions about the relationship between schools, teachers, and students. First, the legislation assumes that clear, coherent standards regarding what students should know will focus school and teacher efforts. Second, NCLB assumes that attaching rewards and sanctions to tests that measure these standards will motivate schools and teachers to work harder. Finally, NCLB assumes that schools will use their resources more wisely and/or apply greater resources to better meet these standards.

In this paper, I focus on NCLB's third assumption regarding the relationship between school resources and student achievement. I look specifically at the assumptions NCLB makes regarding one key educational resource—teachers. NCLB assumes that schools have the capacity to hire and maintain a staff that meets its teacher quality standards. The legislation presupposes that schools face no barriers, such as a limited supply of new teachers and an aging teacher workforce, in putting together a staff that meets these new guidelines. Furthermore, NCLB presupposes that we can establish teacher quality at the point of hire. The legislation is premised on the theory that paper qualifications—a college degree, full certification, and subject matter competency—can adequately distinguish effective teachers from ineffective ones. NCLB theorizes that if schools raise the bar on teacher quality (by screening potential teachers according to federal mandates and/or by demanding that existing teachers update their qualifications) they will likewise raise student performance.

The following hypotheses test the assumptions NCLB makes between teacher traits and student achievement. They address two related questions: (1) How has NCLB changed the composition of teachers' qualifications (i.e., have schools changed their teacher workforces to be more in line with NCLB standards)? and (2) What effect do these NCLB-defined teacher qualifications have on student achievement?

How has NCLB affected the composition of teachers' qualifications?

H_1: NCLB has motivated schools to make a concerted effort to comply with the law as seen in the changing composition of their teacher workforce according to federal guidelines.

The era of increasing accountability motivates school compliance with NCLB's highly qualified teachers mandate. Under NCLB, schools must (1) measure the extent to which all students (e.g., minority and disadvantaged students) have highly qualified teachers; (2) adopt goals and plans to ensure all teachers are highly qualified; (3) publicly report plans and progress in meeting teacher quality goals; and (4) be fully compliant by 2005-2006. This heightened monitoring compels schools to seek out those teacher traits that meet federal approval while deemphasizing those traits not mentioned by NCLB.

What effect do NCLB-defined teacher qualifications have on student achievement?

H_2: Increasing the proportion of teachers with advanced degrees has no impact on student achievement scores.

NCLB assumes that highly qualified teachers possess at minimum a bachelor's degree. Past research, however, suggests only a weak relationship between the amount of schooling a teacher has and the effectiveness of that teacher. Recall Ferguson's (1991) and Darling-Hammond's (1999) respective findings that advanced degrees matter less than other measures of teacher quality such as teacher verbal skills and certification status. Teachers' advanced degrees may not matter particularly at the elementary level

given the basic knowledge demanded in elementary school curricula.

$H_{3:}$ Increasing the proportion of teachers with full certification has no impact on student achievement scores.

California has not revamped its teacher certification process under NCLB. Indeed, in a nationwide analysis of state certification policies, California ranks among the lowest in rigor (National Council on Teacher Quality 2007). Additionally, findings in the education production literature suggest that even when certification is statistically and positively significant, its effect on student achievement is substantively low. That is, alternatively-certified teachers (those with emergency, provisional, or temporary certification) do just as well as their fully-certified colleagues. Given this research and the fact that the state has not significantly retooled its certification process under NCLB, I speculate that increasing the percentage of teachers with full certification will have no effect on student achievement scores.

$H_{4:}$ Increasing the proportion of new teachers decreases student achievement scores.

The proportion of new teachers lowers student achievement scores even though these teachers are more likely to possess the traits deemed desirable by NCLB than veteran teachers. Research indicates that teacher effectiveness becomes apparent *after* the first year of teaching (Murnane and Phillips 1981). While new teachers may be more invested in their students performing well on standardized tests because they entered the profession during the high-stakes regime, they lack the on-the-ground classroom experience necessary to improve student performance.

$H_{5:}$ Increasing the average years teaching experience has no effect on student achievement scores.

Even though veteran teachers may be more effective than new teachers, the returns to their experience levels off. Rivkin *et al.* (2005) found that teacher effectiveness levels off *within* the first three years of teaching. This ceiling effect of teacher experience is even more apparent under the NCLB regime as schools increasingly rely on indicators of teacher quality at the point of hire (i.e., paper qualifications) and place less emphasis on enhancing teacher quality via on-the-job, professional training. Stated differently, there is little payoff to experience as NCLB deemphasizes upgrading the quality of the existing teaching force via professional development and instead stresses the importance of hiring a new crop of teachers that meets its standards.

Data and Methods

Research Site: Los Angeles Unified School District Elementary Schools

I seek to understand the link between NCLB, teacher quality, and student achievement. More specifically, I ask, what effect do teachers who NCLB defines as highly qualified have on student test scores? This type of study requires data on a large number of schools through time. As such, my sample includes all elementary schools (N = 510) within the Los Angeles Unified School District (LAUSD) spanning the period 1999-2006.

I use the above sampling strategy for four reasons. First, I focus my analysis on the school level to avoid the problem of reverse causation. Classroom analyses assume that teachers are randomly assigned to classes. By estimating the relationship between teacher characteristics and student achievement at the school level, I avoid confusing causal relations. Second, I focus my investigation on elementary schools because these

schools spend the bulk of their instructional time on English language arts and math, subjects that are heavily tested under NCLB. Because elementary schools do not have the same curricular flexibility as middle and secondary schools, they are more likely to experience intense NCLB pressures. Thus, NCLB's impact on both teacher workforce and student performance should be more evident at the elementary level. Third, I restrict my study to schools within a single district to control for district-specific administrative policies that may affect student performance, such as funding allocations, classroom size, and teacher retention, assignment, and recruitment procedures. Finally, I confine my study to LAUSD to obtain a varied and sizeable sample of teachers, students, and schools. Historically the largest school district in California, LAUSD enrolls 727,319 students or 12 percent of the state's total student enrollment. According to 2005-2006 data, 18,193 full-time equivalent teachers taught 350,927 students in 510 LAUSD elementary schools.[7]

Test scores as measures of student learning

While academic achievement may be demonstrated in various ways, I focus on standardized test results as policymakers have historically relied on them to evaluate the effectiveness of school reform. As discussed earlier, NCLB is rooted in the high-stakes accountability and standards-based reform movements. These two movements heavily emphasize student learning as measured by performance on tests. By using standardized test scores as a proxy for student learning, I ask, can NCLB meet its own criteria for student academic success?

[7] 2005-2006, California Department of Education Educational Demographics Office.

Time-frame

I examine the time period 1999-2006 to chart the relationship between teacher demographics and student performance before and after NCLB implementation. Data from the years 1999-2000 and 2000-2001 reveal the relationship between teacher demographics and student performance prior to NCLB. Data from 2001-2002, 2002-2003, and 2003-2004 measure this relationship during the initial stages of NCLB implementation. Finally, data from the years 2004-2005 and 2005-2006 tracks changes in the teacher workforce and student achievement after the initial stages of NCLB.

Data Sources and Variables

I collected data from the California Department of Education's integrated accountability data system and basic educational data system. California's integrated accountability data system contains standardized test results for all the state's public schools. The basic educational data system contains credentialed-personnel, student, and program data at the school and district levels.

Academic Performance Index: California's measure of student performance at the school-level

Every school in California receives an annual score called the Academic Performance Index (API) score. The API score is a single number on a scale of 200 to 1,000. The state calculates this score using two assessments. The first, the California Standards Test, constitutes a majority of the API and is based on state-specific academic content standards in English language arts (i.e., reading, language, and spelling) and math. The second testing element that makes up the API is a national norm-referenced test, which also evaluates students in reading, language, spelling, and math. The state

uses this assessment to compare California students to a national sample of students tested in the same grade at the same time of the school year.[8]

Each school's API score is a weighted average of student scores across the different content areas (i.e., English language arts and math) of the California Standards Test and the state's norm-referenced test. Test weights determine how much emphasis each content area has for a particular school's API. These test weights may be calculated differently from year to year. For example, Table 1 shows how the state changed API calculations from 2001-2002 to 2002-2003.

(Table 1 about here)

API Growth: tracking annual changes in student performance at the school-level

Because the state has used various assessments and weighting schemes to calculate the API, this composite score is an unreliable measure of student achievement from one year to the next. However, the growth in a school's API can be used to determine student achievement gains from year to year. This score—the Academic Performance Index Growth (API Growth) score—is calculated using a scale calibration factor, which makes either a positive or negative adjustment to each school's API each year to maintain consistency from one API reporting cycle to the next. Stated differently, the state calculates API Growth scores using the previous year's weighting schemes. Because API Growth more accurately captures annual gains in school performance, I use this as my dependent variable in my statistical analyses.

While more reliable than the API, my dependent variable, API Growth, may still not accurately reflect how much students learn from year to year. This limitation of the

[8] California's norm-referenced test was the Stanford-9 from 1998 until the CAT/6, the California Achievement Text, Sixth edition, replaced it in 2003.

data is common to analyses that seek to chart student growth via standardized tests. As policymakers, educators, and parents have changed their notions of the standards and content students should master so too have standardized tests changed. This is a weakness of all longitudinal studies that rely on such measures. Unfortunately, given the design of the state's testing system, I am restricted to this variable as a measure of annual student growth.

Still, there are compelling reasons for using results from the state's standardized testing system. As mentioned before, by using standardized test scores as a proxy for student learning, I question whether NCLB can meet its own criteria for student academic success. As part of the high-stakes accountability and standards-based reform movements, NCLB is grounded in the theory that students and teachers have to know what is expected of them (i.e., knowledge that is taught in the classroom) if they are to succeed. The API Growth score is calculated from tests that are aligned to the state's curricular standards. This measure captures what California policymakers consider important for students to know and be able to do. Furthermore, California requires that all its teachers be familiar with state-specific content standards. Unlike results from other assessments, which may or may not test knowledge that is taught in California's classrooms, API Growth meets the standards-based movement requirement that knowledge and testing be aligned. As such, API Growth is an appropriate way to evaluate NCLB's theory of action.

API Growth is also an ideal measure of academic achievement because it is based on a rigorous testing system. Under NCLB, states determine test rigor and proficiency levels with the goal of meeting the benchmarks outlined in the federal NAEP (National

Assessment of Educational Progress) tests.[9] Unlike other state testing systems, such as those in Texas and Kentucky, which set the bar low for student achievement, California is recognized for its high testing rigor. For example, the gulf between California and NAEP proficiency standards for 4th grade reading in 2003 is 18 points compared to 58 for Texas and 31 for Kentucky (Mintrop and Trujillo 2005). The rigor of California's testing system ensures that reported student academic achievement is comparable to NAEP, which many scholars consider the gold standard for measuring student learning (Fuller *et al.* 2006; Lee 2006).

Student and teacher demographics

The other variables used in my analysis, retrieved from the state's basic educational data system, include school-level measures of student and teacher demographics. Table 2 lists descriptive statistics and information on the coding of all variables.

(Table 2 about here)

Within the sociology of education literature, researchers have demonstrated the importance of family background in determining student academic achievement. I measure student family background by including the following variables measured at the school level: percentage of students who are poor (those who receive free or reduced-price lunches), English language learners, Asian, black, Hispanic/Latino, and other. The omitted race category is the percentage of students who are white. My measure of parental educational attainment is the percentage of students with parents who have a

[9] First administered in 1969, NAEP is a federally mandated, nationally representative survey of student academic achievement. It tests students in reading, writing, science, math, history, geography, and the arts.

high school degree or less. The omitted category is the percentage of students with parents who have some college or more.

Teacher demographic variables include a measure of teacher certification (the percentage of teachers who hold a full credential), teacher experience (the percentage of teachers who are new and the average years teaching experience for all teachers), and teacher education (the percentage of teachers who have an advanced degree; the omitted category is bachelor's degree). NCLB mandates that all teachers must have at least a college degree on the assumption that some minimum amount of teacher schooling is necessary for classroom effectiveness. However, this distinction is meaningless in the California context as all teacher candidates are required to have at least a bachelor's degree. Instead, I measure the percentage of teachers with advanced degrees to test whether additional years of schooling beyond the minimum NCLB requirements improve student achievement scores.

My measures of teacher quality leave out several traits that other education production scholars have found significant. For example, I have no measure of teacher academic or verbal ability. I also lack indicators of more hard-to-measure teacher characteristics such as teacher dedication, expectation of students, and classroom control. If I were making a more general argument about the relationship between teacher inputs and student outputs, then omission of these other teacher traits would seriously bias my study. However, because I aim to understand whether NCLB's definition of highly qualified teachers promotes student learning, I require only those measures the law deems important. I have two of those measures—teacher certification and teacher education. However, I lack a measure of teacher subject matter competency. The state only began

collecting these data beginning 2005-2006. While this is a potential weakness in my study, I argue that such an omission at the elementary school level is less serious than if my study were focused on secondary schools. That is, elementary school teachers do not require the level of subject expertise that would be needed at the secondary level (Goldhaber and Brewer 2000; Monk 1994). Darling-Hammond (1999) notes that "knowledge of the material to be taught is essential to good teaching, but also the returns to subject mater expertise would grow smaller beyond some minimal essential level which exceeds the demands of the curriculum being taught" (7).

Cross-sectional times series analyses

My dataset has 3,570 observations, which reflect the 510 elementary schools for the seven years (1999-2006).[10] I use cross-sectional time series analyses to examine how teacher quality affects student progress under NCLB. An important advantage of this method is that it allows researchers to investigate the causal relationships in non-experimental studies. Thus, by using a fixed effects regression model, I can study changes in the dependent variable over time while also eliminating the effect of omitted variables that differ across schools but are constant over time.

To determine the effects of NCLB-defined teacher traits on student achievement, I test the following model:

$$\Delta API\ Growth\ score_{i,t+1} = \beta_0 + \beta_1 API\ Growth\ score_{i,t-1} + \beta_2 z_{it} + v_i + \varepsilon_{it},$$

where i indexes the 510 schools and t denotes the 7 years from 1999-2006; $z_{it}=x_t-x_{t-1}$; x is a vector of teacher characteristics and student traits measured at the school level; v_i is the school-specific time-constant error, and ε_{it} is the school-specific and time-varying error.

[10] However, N=2,556 in my regression sample. Some schools could not be included because of missing data on one or more of my variables of interest.

The estimates of the coefficient β_2 show how much the API Growth score changes when schools change from one value to another in an independent variable.

Results

Before examining NCLB's impact on the teacher workforce, it is useful to look first at the general pattern of student achievement pre- and post-NCLB implementation. Figure 1 tracks mean API Growth scores for all LAUSD elementary schools over the time period 1999-2006. The figure reveals that mean API Growth scores have dropped under the NCLB regime. While mean API Growth scores remained relatively stable from 1999-2003 (43.1 in 1999-2000; 38.9 in 2000-2001; 41.0 in 2001-2002; and 41.8 in 2002-2003), a 34 point drop to 7.8 occurred in 2003-2004. In the following year, 2004-2005, the mean API Growth score more than doubled to 16.2 and slightly dropped once more to 14.6 in 2005-2006.[11] The data thus show that schools maintained steady growth in student learning prior to NCLB and lost some ground after the policy's implementation.

(Figure 1 about here)

Teacher workforce trends

To understand the relationship between teacher quality and student achievement, it is also useful to look at the teacher workforce composition before and after NCLB implementation. Establishing these trends reveals whether and how schools responded to

[11] I interviewed Esther Wong, LAUSD's Assistant Superintendent of Planning, Assessment, and Research, about the drop in API Growth scores. She reported that the district initially thought it was due to cohort changes in the third grade but later found that was not the case. My own research confirms that this trend is not unique to LAUSD as API Growth scores dropped across a majority of schools in the state. The evidence suggests that the state altered the test in some way as scores dropped across many of California's schools. In press releases, the California Department of Education acknowledged but did not explain the drop in scores. This highlights a limitation in my modeling approach. By relying on standardized test scores, my model generates an imperfect measure of how school inputs (in this case, teacher qualifications) affect student learning. However, as stated earlier, there are compelling reasons for using the state's standardized test scores as a proxy for student achievement.

NCLB's highly qualified teacher mandate. Figure 2 shows the percentage distribution of teachers with full certification, teachers with graduate degrees, and new teachers for all LAUSD elementary schools over the time period 1999-2006. The data show that schools "upgraded" the quality of their teaching staff over the entire time period, with the most rapid changes occurring immediately after NCLB implementation (2002-2003). Prior to NCLB (1999-2000), the mean percentage distribution of fully-credentialed teachers in LAUSD elementary schools was 71.6 percent. Starting with the initial stages of NCLB implementation, this number rises 15.8 percentage points, from 79.2 percent in 2002-2003 to 95.0 percent in 2005-2006. Schools thus made a concerted effort to fill their teaching staff with fully-credentialed personnel following NCLB's passage.

Figure 2 also shows that the percentage of teachers with graduate degrees also increased following NCLB implementation although this change was less dramatic than changes seen in the number of fully-credentialed teachers. The mean percentage distribution of teachers with graduate degrees prior to NCLB was 24.0 percent in 1999-2000. Five years after the introduction of NCLB (2005-2006), the percentage of teachers with graduate degrees increases 8 percentage points to 32.0 percent.

(Figure 2 about here)

Additionally, LAUSD elementary schools sought more experienced teachers following NCLB's enactment by decreasing the proportion of new teachers. Figure 2 shows that the percentage of new teachers reduces by more than a third from 2001-2002 (18.0 percent) to 2005-2006 (10.8 percent). As the proportion of new teachers decreases over the time period 1999-2006, the average years of teaching experience in the district

rises (Figure 3). During this time, the average years of teaching experience for all teachers to a high of 11.4 years from a low of 10.8 years (2000-2001).

(Figure 3 about here)

Although the data in Figure 4 reflect the number of new teacher hires at both the elementary and secondary level and only for the time period 2000-2005, it nevertheless provides interesting, additional information on NCLB's impact on the teacher workforce. The figure shows that the district focused more intensely on hiring individuals that meet NCLB highly qualified status during the initial stages of NCLB implementation (2001-2004) than later (2004-2005). The percentage of new teacher hires that are both fully-credentialed and subject-matter competent more than doubles during the initial period of NCLB implementation, from 41.5 percent in 2001-2002 to 92.6 percent in 2003-2004. Conversely, traits not deemed important by NCLB (i.e., pre-intern teachers holding emergency credentials) drops by 51 percentage points during this same initial period (from 58.5 percent in 2001-2002 to 7.0 percent in 2004-2005). Schools thus sought a new cohort of teachers that met NCLB's fully-credentialed and subject-competent criteria.

(Figure 4 about here)

The data clearly show that schools responded to NCLB's highly qualified teacher mandate and heightened monitoring by changing the composition of their teaching staff to meet the policy's standards. Schools executed most of these changes during the initial years of NCLB enactment. Over the time period 1999-2006, the district increased its proportion of teachers with full certification and advanced degrees. Additionally, even though NCLB is silent on the issue of new vs. experienced teachers, schools reduced the

number of new teacher hires and more successfully retained their more experienced teachers. Among the new teachers hired, an overwhelming majority satisfies all three components of NCLB's highly qualified teachers mandate. Indeed, LAUSD's teacher workforce composition changed significantly following the enactment of NCLB.

Impact of teacher workforce on school performance

Results from my regression analysis reveal the relationship between NCLB-defined teacher qualifications and student achievement. The six student variables that have statistically significant effects on a school's API Growth score include the previous year's API Growth score; the change in the percentage of students who are poor; the change in the percentage of students who are English language learners; the change in the percentage of students who are black; the change in the percentage of students who are Hispanic/Latino; and the change in the percentage of parents who have a high school degree or less.

A school's API Growth score from the previous year has a negative effect (-1.01) on its API Growth score for the following year. We would expect this as schools that grow considerably the previous year would show smaller gains the following year. Conversely, we would expect schools that post negative API Growth scores the previous years to increase their API Growth score the following year as underperforming schools have more room to improve. Put another way, a school's prior achievement significantly determines its current achievement.

Similarly, increasing the percentage of students who are black or Hispanic/Latino decreases API Growth scores. The magnitude of these effects are comparable: a one percent increase in the proportion of students who are black decreases the API Growth

score by -1.61 points while a one percent increase in the proportion of students who are Hispanic/Latino decreases the API Growth score by -1.24. Thus, if a school experienced a 12 percent increase in the proportion of its black students (the maximum for my sample), then its API Growth score for the following year would decline 10 points. Likewise, if a school experienced a 12 percent increase in the proportion of its Hispanic/Latino students (the maximum for my sample), then its API Growth score for the following year would decline by 10 points. This finding is consistent with the literature on the achievement gap, which show that blacks and Hispanics/Latinos consistently underperform in the school setting.[12]

Surprisingly, the data show that changes in the percentage of poor students, English language learners, and parents with high school degree or less are positively associated with changes in the API Growth score. However, when compared to the magnitudes of changing the percentage of black or Hispanic/Latino students, the magnitude of these effects— .57 for poor students, .25 for English language learners, and .13 for parents with a high school degree or less—is small. This finding that poverty, language status, and parental education is significantly and positively correlated to API Growth scores suggests that schools may more easily overcome the barriers associated with student socioeconomic status than race.

Only two of the four teacher traits—changes in the percentage of both new teachers and teachers with graduate degrees—have statistically significant effects on changes in the API Growth score. Consistent with my hypothesis, increasing the percentage of new teachers decreases API Growth scores. The magnitude of this effect is

[12] See Jencks and Phillips (1998) for an extensive collection of articles that discuss the black-white test score gap.

small; a one percent increase in the proportion of new teachers decreases API Growth scores by less than half a point (-.31). In my sample, schools on average annually decreased their proportion of new teachers by 2.03 percent; this change translates into a very small gain in the API Growth score, merely 1 point.

Contra my hypothesis, changing the percentage of teachers with advanced degrees *does* have an impact on API Growth score gains. The magnitude of changing the proportion of teachers with advanced degrees is almost twice that associated with changes in the proportion of new teachers (-.60 compared to -.31). Thus, teachers with graduate degrees have a *larger* negative impact on API Growth scores than teachers who are new to the profession. However, the magnitude of this effect is relatively small. In my sample, schools on average annually increased the percentage of teachers with graduate degrees by 1.29 percent; this change results in a 1 point decline in API Growth scores.

Consistent with my hypotheses, changes in the percentage of teachers with full certification and changes in the average years teaching experience do *not* have statistically significant effects on changes in the API Growth score.

The data clearly show that student traits have a larger effect on changes in API Growth scores than teacher traits. Six of the eight student variables measured have statistically significant effects on changes in API Growth score compared to only two of the four teacher traits. Moreover, the magnitude of the effects of all of the student variables—with the exception of the percentage of students who are English language learners and the percentage of parents who have a high school degree or less—is larger than the magnitude of the teacher effects.

These findings, however, do not suggest that teachers are irrelevant in the educational process. Rather, the results question whether the paper qualifications that NCLB uses to evaluate teachers do in fact distinguish effective teachers from ineffective ones. The data does not support NCLB's assumptions regarding the amount of schooling that effective teachers possess. In fact, the results show that increasing the proportion of teachers with advanced degrees actually drives down API Growth scores. Thus, NCLB's mandate that teachers have at minimum a bachelor's degree may be too crude to accurately determine teacher quality. More detailed information about the type of degree (e.g., whether the degree is in the subject(s) taught and the amount of coursework involved) and which institution conferred that degree (i.e., college reputation) may be needed to separate effective teachers from ineffective ones.

The data also cast doubt on NCLB's mandate that all teachers be fully certified. Changing the percentage of teachers with full certification has no statistically significant impact on changes in the API Growth score. As it stands now, NCLB is silent on what constitutes a high-quality teacher certification program and does not recognize that the value of full certification may vary from state to state. Teachers with full certification may in fact be more effective than teachers with alternative certification but only if that full certification is obtained through a rigorous process. A California certification may have no impact on student achievement because, as mentioned before, the state's certification standards compares poorly to other states'.

Finally, the data highlights that other teacher variables not included in NCLB's definition should be taken into account when determining teacher quality. NCLB is silent on the issue of new vs. experienced teachers yet my results indicate that new teachers

negatively affect student performance. These new teachers may lack the on-the-job experience critical for classroom effectiveness. Interestingly, the new teachers in my sample are also more likely to possess all of the traits that NCLB considers important (see Figure 4).

(Table 3 about here)

Conclusions

Schools, teachers, and students have faced increasing scrutiny under the NCLB regime. Student learning is now heavily monitored via annual standardized testing. The rationale behind this heightened monitoring traces back to longstanding public dismay over poor student performance. Today's policymakers and educational reformers seek to reverse the decline in student achievement by attaching high stakes to student progress. Under NCLB, schools risk losing funding and may be labeled "failing" for student underperformance.

Attaching rewards and sanctions is not the only way NCLB attempts to stimulate student academic growth. The legislation has also put forth an unprecedented set of federal guidelines for teacher quality in an effort to raise student learning. NCLB theorizes that student achievement will improve if schools are held accountable for teacher quality standards. Prior to NCLB, collegiate schools of education, school administrators, teachers, and parents decided the measures of teacher quality. NCLB marks the federal government's first serious intervention in matters of teacher quality. Just as the high-stakes accountability and standards-based movement seek to increasingly regulate student learning, NCLB seeks greater oversight of the teacher workforce.

The data show clearly that NCLB has succeeded in changing the composition of teachers' qualifications. Beginning with the law's enactment, schools altered their teacher workforce to more closely reflect NCLB standards of teacher quality. Schools both maintained and recruited greater proportions of teachers with full-credentials, advanced degrees, and subject matter competency. Faced with increasing federal regulation, they altered their teaching staffs despite uncertain returns to student achievement.

Indeed, the results of this analysis cast doubt on whether NCLB-defined teacher characteristics translate into classroom effectiveness. My findings encourage us to revisit NCLB's assumption that teacher quality is best determined at the point of hire. It questions initiatives that focus on hiring a new crop of teachers via paper-based qualifications such as a college degree and full certification.

Contra NCLB's assumptions, I found a negative relationship between increasing the proportion of teachers with additional years of schooling and student growth. Teachers with master's and doctorate degrees actually drive down a school's API Growth score. This counterintuitive finding suggests that crude distinctions between different types of educational degrees (i.e., bachelor's, master's, and doctorate) may not adequately distinguish effective teachers from ineffective ones. Rather, as other scholars have suggested, a more meaningful distinction would take into account whether the degree was in the subject(s) taught and the reputation of the institution that conferred the degree.

Moreover, the data caution against the value of full certification as a measure of teacher quality. I found no statistically meaningful relationship between teachers who are

fully certified and student achievement. This is not to suggest, however, that a teachers' certification status has no predictive power with regards to student achievement. Rather, the effect of certification status may depend on the rigor and quality of the teacher's certification program. NCLB's teacher mandate currently does not recognize this distinction.

Finally, the finding that new teachers negatively affect student achievement scores suggests that other methods of raising teacher quality deserve further consideration. Ironically, an overwhelming majority of the new teachers in my sample meet all three of NCLB's teacher standards. The data thus suggests that investing in better training and mentoring for new teachers could be a fruitful way to raise the quality of the teacher workforce. Instead of focusing merely on establishing teacher quality at the point of hire, NCLB could invest more in the professional development of both new and veteran teachers. Just as individuals in other professions become better at their job through experience and ongoing training, teachers may become more effective with classroom experience and meaningful professional development. This alternative approach towards raising teacher quality is worth further exploration.

References

Ballou, Dale, and Michael Podgursky. 1998. "The case against teacher certification." *Public Interest* 132:17-29.

Ballou, Dale, and Michael Podgursky. 2000. "Reforming teacher preparation and licensing: What is the evidence?" *Teachers College Record* 102:5-27.

Betts, Julian R., Kim S. Rueben, and Anne Danenberg. 2000. *Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California*: Public Policy Institute of California.

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2006. "How Changes in the Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." *Education Finance and Policy* 1:176-216.

Coleman, James S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, and R. L. York. 1966. "Equality of Educational Opportunity." Washington, DC: U.S. Government Printing Office.

Darling-Hammond, Linda. 1999. "Teacher Quality and Student Achievement: A Review of State Policy Evidence." Center for the Study of Teaching and Policy.

Darling-Hammond, Linda, Deborah J. Holtzman, Su Jin Gatlin, and Julian Vasquez Heilig. 2005. "Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness." *Education Policy Analysis Archives* 13(42):1-48.

Ehrenberg, Ronald G., and Dominic J. Brewer. 1994. "Do School and Teacher Characterisitics matter?: Evidence from High School and Beyond." *Economics of Education REview* 13:1-17.

Ferguson, Ronald F. 1991. "Paying for Public Education: New Evidence on How and Why Money Matters." *Harvard Journal on Legislation* 28:465-498.

Fuller, Bruce, Kathryn Gesicki, Erin Kang, and Joseph Wright. 2006. "Is the No Child Left Behind Act Working? The Reliability of How States Track Achievement." Pp. 1-43: Policy Analysis for California Education.

Goldhaber, Dan D., and Dominic J. Brewer. 1997. "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." *Journal of Human Resources* 32:505-523.

Goldhaber, Dan D., and Dominic J. Brewer. 2000. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis* 22:129-145.

Gordon, Robert, Thomas. J. Kain, and Douglas O. Staiger. 2006. *Identifying effective teachers using performance on the job: The Hamilton Project (Discussion Paper 2006-01)*. Washington, DC: The Brookings Institution.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66:361-396.

Hamilton, Laura S., Brian M. Stecher, and Stephen P. Klein (Eds.). 2002. *Making Sense of Test-Based Accountability in Education*. Santa Monica: RAND.

Hanushek, Eric. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *The American Economic Review* 61:280-288.

Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24:1141-1177.

—. 1997. "Assessing the effects of school resources on student performance: An Update." *Educational Evaluation and Policy Analysis* 19:141-164.

—. 2003. "The failure of input-based schooling policies." *Economic Journal* 113:F64-F98.

Hanushek, Eric A., and Steven G. Rivkin. 2003. "How to Improve the Supply of High Quality Teachers." in *Brooking Papers on Education Policy*: Brookings.

Heubert, Jay P., and Robert M. Hauser. 1999. "High Stakes: Testing for Tracking, Promotion, and Graduation." National Research Council.

Hirsh, Deborah. 2005. "Teacher Quality at LAUSD: A Continuing Success Story." Los Angeles Unified School District.

Jencks, Christopher, and Meredith Phillips (Eds.). 1998. *The Black-White Test Score Gap*. Washington, D.C.: Brookings Institution Press.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2005. "Identifying Effective Teachers in New York City." NBER Summer Institute.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2007. "Photo Finish: Certification Doesn't Guarantee a Winner." *Education Next* 7:60-67.

Lee, Jaekyung. 2006. "Tracking Achievement Gaps and Assessing the Impact of NCLB on the Gaps: An In-depth Look into National and State Reading and Math Outcome Trends." Cambridge: The Civil Rights Project at Harvard University.

Mintrop, Heinrich, and Tina Trujillo. 2005. "Corrective Action in Low-Performing Schools: Lessons for NCLB Implementation from State and District Strategies in First-Generation

Accountability Systems." in *Annual Conference of the American Educational Research Association*. Montreal.

Monk, David H. 1994. "Subject area preparations of secondary mathematics and science teachers and student achievement." *Economics of Education Review* 13:125-145.

Murnane, Richard, and Barbara R. Phillips. 1981. "What do effective teachers of inner-city children have in common?" *Social Science Research* 10:83-100.

National Commission on Excellence in Education. 1983. "A Nation at Risk: The Imperative for Educational Reform." United States Department of Education.

National Council on Teacher Quality. 2007. "California State Summary." in *State Teacher Policy Yearbook: Progress on Teacher Quality*.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. "How Large are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26:237-57.

Pipho, Chris. 1985. "Tracking the Reforms, Part 5: Testing—Can it Measure the Success of the Reform Movement?" Education Week 4:19.

Ravitch, Diane. 2002. "Testing and Accountability, Historically Considered." in School Accountability, edited by W. M. Evers and H. J. Walberg. Stanford: Hoover Institution Press.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2001. "Teachers, schools, and academic achievement." National Bureau of Economic Research.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73:417-458.

Smith, Marshall. S., Jennifer. O'Day, and David K. Cohen. 1990. "National Curriculum American Style: Can It Be Done? What Might It Look Like?" *American Educator* 14:10-17, 40-47.

Sunderman, Gail L., James S. Kim, and Gary Orfield. 2005. *NCLB Meets School Realities: Lessons From the Field*. Thousand Oaks: Corwin Press.

Summers, Anita A., and Barbara L. Wolfe. 1977. "Do Schools Make a Difference?" *The American Economic Review* 67:639-652.

United States Government Accountability Office. 2007. "Teacher Quality: Approaches, Implementation, and Evaluation of Key Federal Efforts." Subcommittee on

Higher Education, Lifelong Learning, and Competitiveness, Committee on Education and Labor.

Walsh, Kate. 2001. "Teacher Certification Reconsidered: Stumbling for Quality." Abell Foundation.

Wayne, Andrew J., and Peter Youngs. 2003. "Teacher Characteristics and Student Achievement Gains: A Review." *Review of Educational Research* 73:89-122.

West, Martin R., and Paul E. Peterson. 2003. "The Politics and Practice of Accountability." in *No Child Left Behind? The Politics and Practice of School Accountability*, edited by Paul E. Peterson and Martin R. West. Washington, D.C.: Brookings Institution.

Appendix A

Table A1. California's NCLB Teacher Requirements: Elementary School Teachers

"New" to the Profession Holds a Credential or an Intern Credential or Certificate issued on or after July 1, 2002	"Not New" to the Profession Holds a Credential or an Intern Credential or Certificate issued before July 1, 2002
1) Bachelor's degree	4) Bachelor's degree
2) California Credential or an Intern Credential or Certificate for no more than three years	5) California Credential or an Intern Credential or Certificate for no more than three years
3) Core academic subject competence must be demonstrated by: EXAM: Pass a multiple subjects examination approved by the California Commission on Teacher Credentialing (CCTC)	6) Core academic subject competence must be demonstrated by: EXAM: Pass a multiple subjects examination approved by the California Commission on Teacher Credentialing (CCTC) OR HOUSSE: Complete California's High Objective Uniform State Standard of Evaluation; "Not new" teachers may demonstrate core academic subject-matter competence in multiple ways though a combination of: • Prior experience in the core academic content area • Course work in the core academic content area • Standards-aligned professional development in the core academic content area • Leadership and service to the profession in the core academic content area • Observation and portfolio assessment in the core academic content

Source: California Department of Education.

Appendix A

The No Child Left Behind Act

The No Child Left Behind Act (NCLB), the legislation that mandates high-stakes testing, has set the agenda for local school reform since its passage in 2001. Under the law, schools, not students, are held accountable for low student achievement. Standards, testing, and rewards/sanctions make up the law's three central elements.

NCLB requires that states put together a set of standards and design assessments to measure whether those standards are being met. To receive Title I federal funding, states must test students annually in grades 3-8 and 10-12.[13] Schools must also release results to the public and demonstrate that students are making adequate yearly progress toward full proficiency. NCLB allows states to specify the amount of progress required each year. However, it mandates that all states establish a single minimum proficiency rate that applies to all schools as well as to all subgroups of students (broken out by poverty, race, ethnicity, disability, and limited English proficiency) within schools. Once a state defines proficiency, the minimum proficiency rate for each school and subgroup is defined as the maximum of the proficiency rate of the 20th percentile school or the proficiency rate of the lowest scoring subgroup. The minimum proficiency level must increase at regular intervals until it reaches 100 percent at the end of 12 years.

The law determines what sanctions apply and when. A failing school is one where any of the subgroups within the school fails to achieve the minimum proficiency rate. Schools that fail to improve after two consecutive years are identified as in need of improvement and enter into Program Improvement (PI) status. These schools must give parents the option of removing their child to another school within the same district (districts must bear the transportation expense). Schools that fail for three consecutive years must offer vouchers to low-income students for use in supplemental educational services such as after-school tutoring programs. After four consecutive years of failure, schools must implement one of several corrective actions, such as implementing a new school curriculum. Schools with a five-year failure record will be restructured (options include overhauling staff; switching over to private administration; conversion into charter school status).

[13] No state has chosen to opt out of NCLB requirements by refusing Title I resources.

Table 1. Changes in test weights used to calculate Academic Performance Index (API) scores in 2001-2002 and 2002-2003.

Content area	2001-2002 API test weight	2002-2003 API test weight
CST[1] English language arts	.36	.48
CST math	.00	.32
NRT[2] reading	.12	.06
NRT language	.06	.03
NRT spelling	.06	.03
NRT math	.40	.08
Total	1.00	1.00

Note: [1]CST=California Standards Test; [2]NRT=Norm-referenced test.
Source: California Department of Education.

Table 2. Descriptive Statistics for Variables Used in the Analysis.

Label	Description	Mean	SD	Min	Max
API Growth score (lag)	School's API growth score, lagged one year.	31.05	27.54	-121.00	129.00
ΔPoor students	Change in the percentage of students tested that participate in the Free or Reduced Price Lunch Program between year t and t-1.	-.66	5.21	-70.00	36.00
ΔEnglish language learners	Change in the percentage of students who are English language learners between year t and t-1.	-.63	4.70	-46.00	16.00
ΔAsians	Change in the percentage of students who are Asian between year t and t-1.	-.03	1.06	-6.00	14.00
ΔBlacks	Change in the percentage of students who are black between year t and t-1.	-.32	1.62	-8.00	12.00
ΔHispanics/Latinos	Change in the percentage of students who are Hispanic/Latino between year t and t-1.	.54	2.23	-17.00	12.00
ΔOther	Change in the percentage of students who are American Indian, Pacific Islander, and Filipino between year t and t-1.	.09	.98	-4.00	9.00
ΔParents high school degree or less	Change in the percentage of parents who have a high school degree or less between year t and t-1.	0.43	13.00	-89.00	100.00
ΔFull credential	Change in percentage of teachers with full credentials between year t and t-1.	4.15	7.84	-89.00	80.00
ΔAverage years teaching	Change in the average number of total years that teachers have been instructing between year t and t-1.	.13	1.19	-9.27	9.02
ΔNew teachers	Change in percentage of first- and second-year teachers between year t and t-1.	-2.03	7.61	-31.82	81.82
ΔTeachers with graduate degree	Change in the percentage of teachers with a master's degree or higher between year t and t-1.	1.28	5.29	-50.00	33.82

Source: California Department of Education integrated accountability data system; California basic educational data system.

Table 3. Regression of Change in API Growth score, All LAUSD Elementary Schools, 1999-2006: Fixed Effects Model.

	Dependent Variable Change in API Growth score		
API Growth score (lag)	-1.014	.022	*
ΔPoor students	.573	.110	*
ΔEnglish language learners	.254	.130	*
ΔAsians	-.375	.550	
ΔBlacks	-1.606	.430	*
ΔHispanics/Latinos	-1.239	.337	*
ΔOther	.528	.585	
ΔParents high school degree or less	.131	.041	*
ΔFull credential	-.140	.081	
ΔAverage years teaching	-.642	.568	
ΔNew teachers	-.311	.083	*
ΔTeachers with graduate degree	-.596	.117	*
Constant	27.370	.995	
Sigma_u	14.328		
Sigma_e	26.214		
N	2,556		

*p<.05

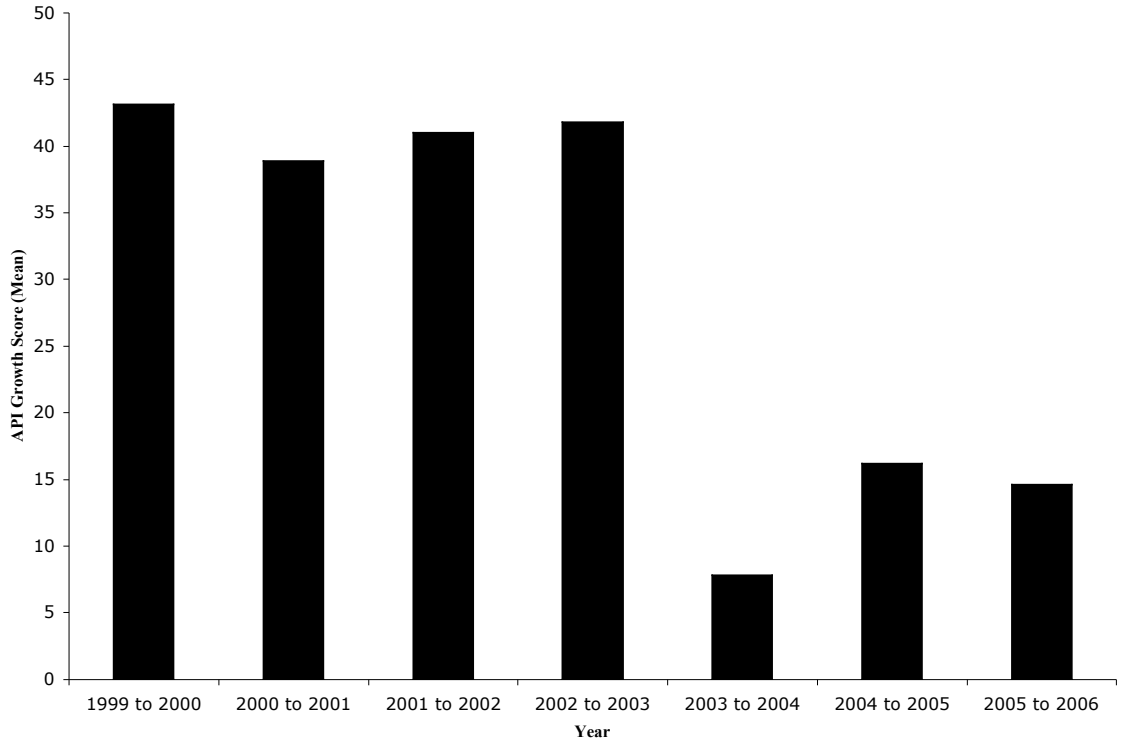

Figure 1. Mean API Growth Score: All LAUSD Elementary Schools, 1999-2006.
Source: California Department of Education.

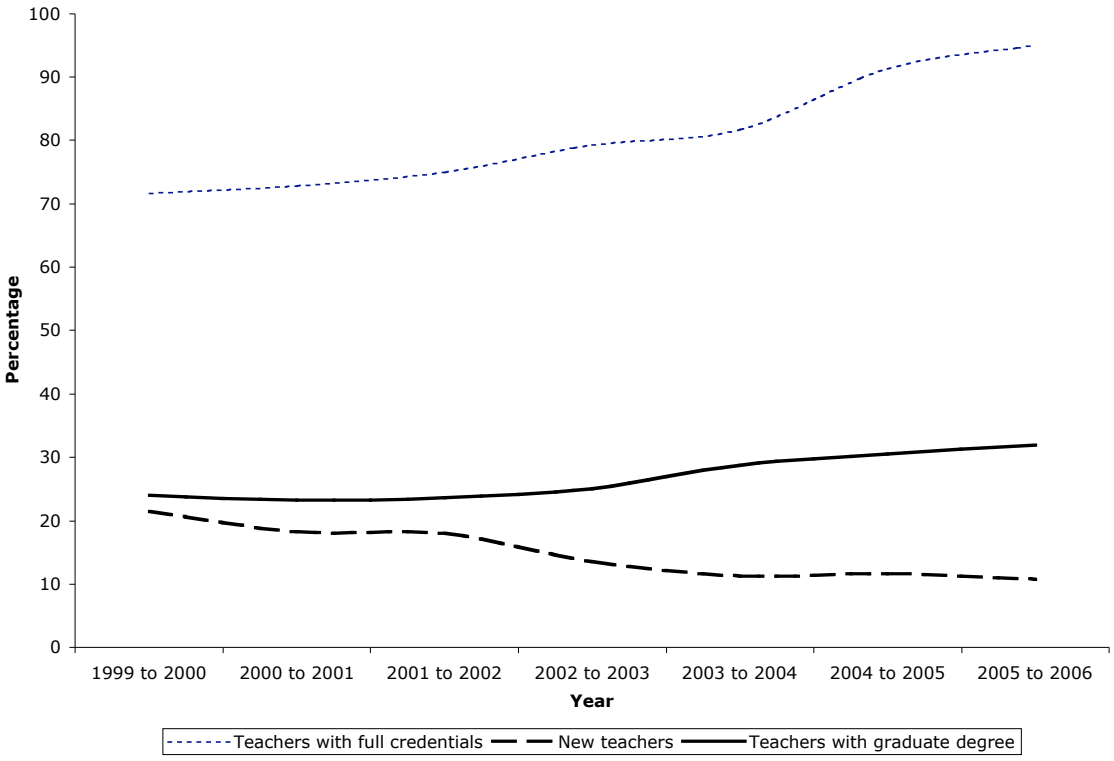

Figure 2: Percentage Distribution of Fully-Credentialed Teachers, New Teachers, and Teachers with Graduate Degrees: All LAUSD Elementary Schools, 1999-2006.
Source: California Department of Education.

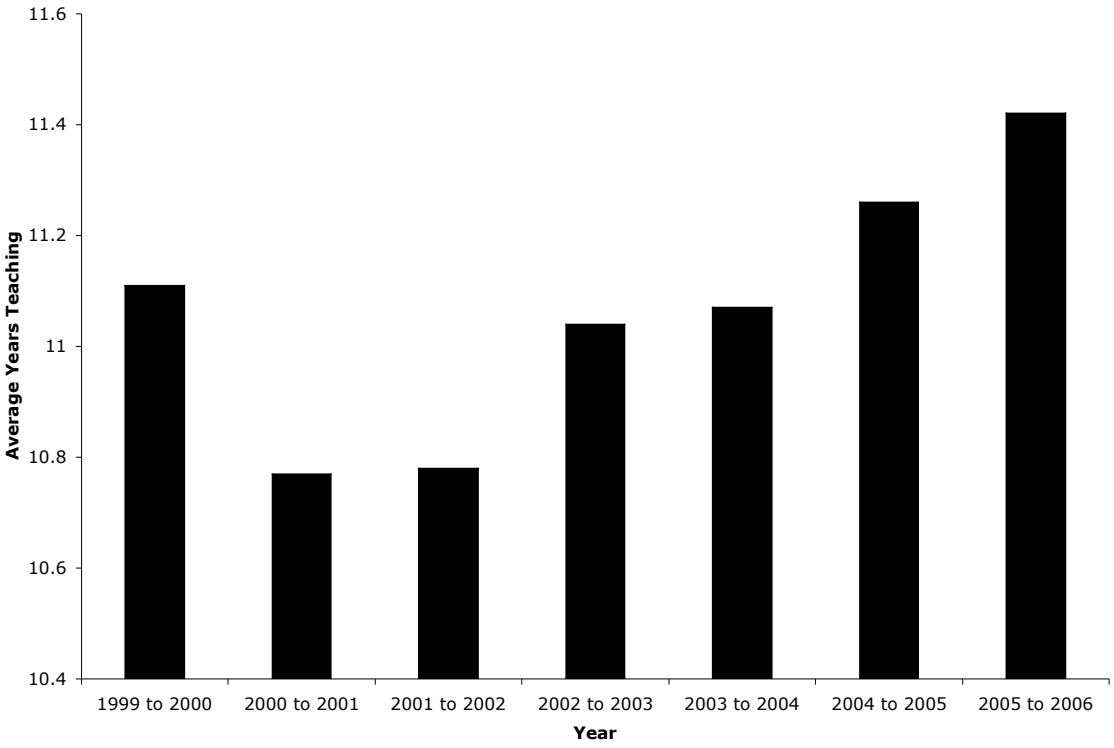

Figure 3: Average Years Teaching Experience: All LAUSD Elementary Schools, 1999-2006.
Source: California Department of Education.

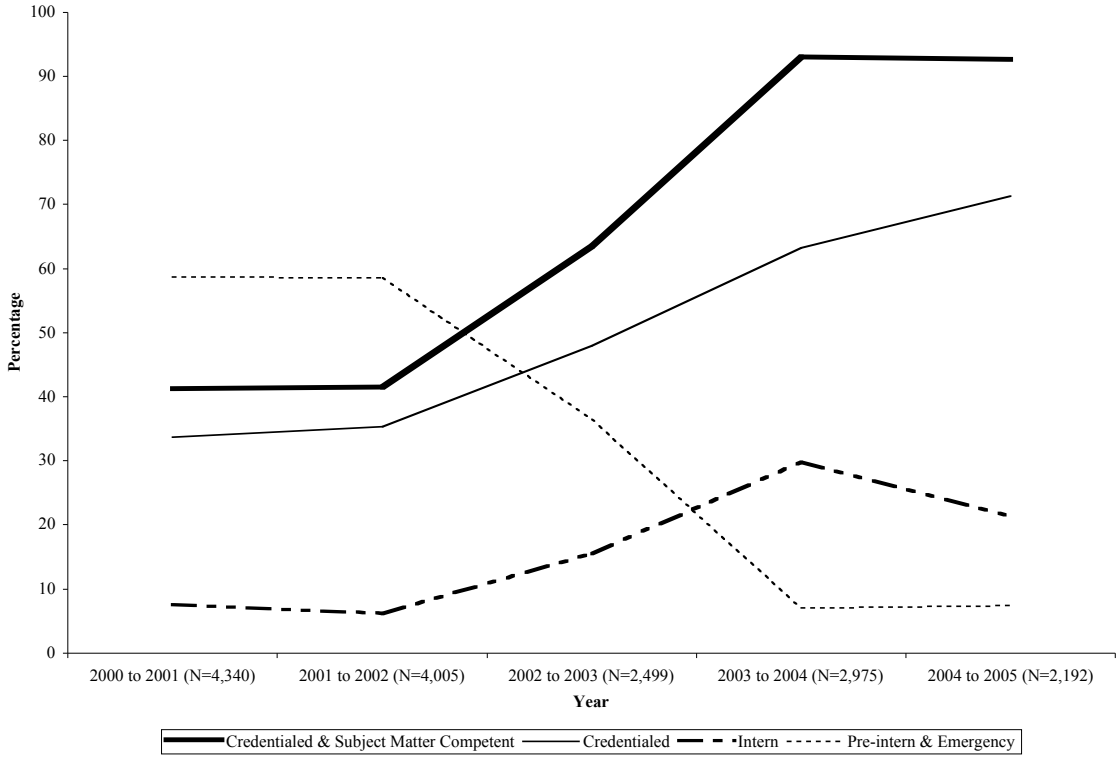

Figure 4. Percentage Distribution of New Teacher Hires by Certification Status: All LAUSD Schools, 2000-2005.
Source: Author's compilation of data in Hirsh (2005:13).