EDUCATION

THE ARTS
CHILD POLICY
CIVIL JUSTICE
EDUCATION
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INTERNATIONAL AFFAIRS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
SUBSTANCE ABUSE
TERRORISM AND HOMELAND SECURITY
TRANSPORTATION AND INFRASTRUCTURE
WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

## Support RAND

Browse Books & Publications

Make a charitable contribution

## For More Information

Visit RAND at www.rand.org

Explore RAND Education

View document details

This product is part of the RAND Corporation reprint series. RAND reprints present previously published journal articles, book chapters, and reports with the permission of the publisher. RAND reprints have been formally reviewed in accordance with the publisher's editorial policy, and are compliant with RAND's rigorous quality assurance standards for quality and objectivity.

# Teacher Qualifications and Student Achievement in Urban Elementary Schools

**Richard Buddin[*] and Gema Zamarro[+]**
**May 2009**

## ABSTRACT

Teacher quality is a key element of student academic success, but few specific teacher characteristics influence classroom outcomes. This research examines whether teacher licensure test scores and other teacher attributes affect elementary student achievement. The results are based on longitudinal student-level data from Los Angeles. California requires three types of teacher licensure tests as part of the teacher certification process; a general knowledge test, a subject area test (single subject for secondary teachers and multiple subject for elementary teachers), and a reading pedagogy test for elementary school teachers. The student achievement analysis uses a value-added approach that adjusts for both student and teacher fixed effects. The results show large differences in teacher quality across the school district, but measured teacher characteristics explain little of the difference. Teacher licensure test scores are unrelated to teacher success in the classroom. Similarly, student achievement is unaffected by whether classroom teachers have advanced degrees. Student achievement increases with teacher experience, but the linkage is weak and largely reflects poor outcomes for teachers during their first year or two in the classroom.

(JEL: J44, J45, H0, H75, I21)
(Keywords: Teacher quality, teacher licensure, student achievement, two-level fixed effects, education production function)

---

Note: The authors are grateful to Harold Himmelfarb of the Institute of Education Sciences for his encouragement and support of this research. We are indebted to David Wright and William Wilson of the California State University (CSU), Office of the Chancellor, for providing access to teacher licensure test score data for recent graduates of the CSU system. Cynthia Lim and Glenn Daley of the Los Angeles Unified School District (LAUSD) provided access to student achievement data and answered numerous questions about district policies and procedures. Eva Pongmanopap of LAUSD was helpful in building the student achievement files and in clarifying numerous issues about the data. Ron Zimmer and Jerry Sollinger provided comments on an earlier draft. We are also grateful to the editor and two reviewers for their comments on an earlier version of this paper. This paper is part of a larger research project "Teacher Licensure Tests and Student Achievement" that is sponsored by the Institute of Education Sciences in the United States Department of Education under grant number R305M040186.

[*] Corresponding author. RAND Corporation, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138, Fax: 310 260-8176, Tel: 310 393-0411 x7875, E-mail: buddin@rand.org
[+] RAND Corporation, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138, E-mail: gzamarro@rand.org

## 1. INTRODUCTION

Urban school districts face special challenges in educating youth. First, these districts often have disproportionate shares of low-income and at-risk students. These students are often concentrated in some schools and neighborhoods, so this isolation means these students have little interaction with more affluent peers. Second, teachers prefer to work near home, so they gravitate towards more affluent suburbs or wealthier neighborhoods in urban districts (Boyd et al., 2005). This sorting means urban districts and schools in low-income urban neighborhoods often employ teachers with low qualifications and weak academic credentials to instruct disproportionate shares of low income and at-risk students (Murnane and Steele, 2007). These poorly prepared teachers have difficulties in the classroom and often leave the teaching profession or transfer to less arduous duty in suburban schools. Finally, if urban schools provide weak preparation for their at-risk students, then these students will have limited opportunities for advancement in a technological economy and the pool of qualified teachers will be further eroded.

These types of concerns have made improving teacher quality (especially in urban areas) a pervasive concern of parents, educators, and policymakers. The concern is driven by the perception of lagging student achievement, especially for at-risk minority students and students from disadvantaged families. In 1998, the Title II (Teacher Quality Enhancement Grants for States and Partnerships) legislation encouraged states to institute mandated teacher testing as part of initial state teacher certification. The No Child Left Behind (NCLB) Act of 2001 required a "highly qualified teacher" in all classrooms and public reporting of teacher qualifications. In addition to the national policies, teacher quality and student achievement progress have been key issues in state and local elections debates throughout the country.

The push for improved teacher quality is being driven by several studies that have shown substantial differences in student achievement across different teachers (Wright et al., 1997;

Rowan et al., 2002; Rivkin et al., 2005). However, the empirical evidence has thus far failed to identify specific teacher characteristics (e.g., experience, professional development, and higher-level degrees) that are linked to higher achievement scores. This mix of results creates a dilemma for educators and policy makers—some teachers are much more successful than others in the classroom, but there is no persuasive evidence on how to raise the overall quality of classroom teaching.

This research examines the relationship between teacher quality and student achievement performance. Specifically, this study addresses three issues.

1. How does teacher quality vary across classrooms and across schools? The analysis uses longitudinally linked student-level data to examine whether students consistently perform better in some teachers' classrooms than in others. The study also assesses whether "high quality" teachers are concentrated in a portion of schools with well-prepared, motivated students or whether higher performing teachers teach both high- and low-performing students.
2. Do traditional measures of teacher quality like experience and teacher educational preparation explain their classroom results? Teacher pay is typically based on teacher experience and education level (Buddin et al., 2007), so it is important to assess whether these teacher inputs are tied to better classroom outcomes.
3. Does teacher success on licensure test exams translate into better student achievement outcomes in a teacher's classroom? Licensure tests restrict entry into teaching (especially for minority teaching candidates), and considerable resources are expended on these exams. In most cases, the cutoff scores for licensure tests are determined by education experts who assess the minimum levels of skill and knowledge "needed" for beginning teachers. But these judgments are not cross-validated by assessing how well these traits subsequently translate into teaching performance in the classroom.

The answers to these types of questions will help policymakers to understand differences in teaching quality and to construct policies and incentives for improving the quality of the teacher workforce.

The study focuses on elementary school students in Los Angeles Unified School District (LAUSD). LAUSD is the second largest school district in the United States with K-12 enrolments of about 730,000 students per year. The data consist of five years of student-level achievement data where individual students are linked to their specific classroom teacher each year. The analysis is based on a sample of over 300,000 students in grades 2 through 5, and these students are taught by over 16,000 different teachers. The longitudinal nature of the data allows us to track student achievement progress of students from year to year in different classrooms and with different teachers. The LAUSD achievement data are augmented with information on teacher licensure test scores for new teachers, as well as more traditional measures of teacher credentials like experience and educational background.

The remainder of the paper is divided into four sections. The second section reviews prior literature on teacher quality and licensure test scores. Several key empirical issues are discussed there which are critical for disentangling how teachers affect student achievement from the types of students assigned to each teacher. The third section describes the econometric approach and database used in the analysis. Section four reports the results. The final section offers conclusions and recommendations.

## 2. LITERATURE REVIEW

### Teacher Quality

Research on teacher effectiveness has progressed through three distinct stages, reflecting data availability and emerging empirical approaches. Initial studies relied on cross sectional data aggregated at the level of schools or even school districts (Hanushek, 1986). This approach related average school test scores to aggregate measures of teacher proficiency. Hanushek (1986)

showed that most explicit measures of teacher qualifications like experience and education had little effect on student achievement. In contrast, implicit measures of teacher quality (i.e., the average performance of individual teachers) differed significantly across teachers. These studies lacked controls for the prior achievement of students attending different groups of schools. If districts assigned teachers with stronger credentials to schools with better prepared students, then the estimated return to teacher credentials would be overstated.

A new round of studies focused on year-to-year improvements in student achievement. These studies implicitly provided better controls for student background and preparation by isolating individual student improvements in achievement. They provided some evidence for differences in teacher qualifications affecting student achievement gains. For example, Ferguson (1991) found that scores on the teacher licensing test in Texas—which measures reading and writing skills as well as a limited body of professional knowledge—accounted for 20-25 percent of the variation across districts in student average test scores, controlling for teachers' experience, student-teacher ratio, and percentage of teachers with master's degrees. Ferguson and Ladd (1996) found smaller effects using ACT scores in Alabama. Ehrenberg and Brewer (1995) found that the teacher test scores on a verbal aptitude test were associated with higher gains in student scores although the results varied by school level and students' racial/ethnic status. Using data from the 1998 National Educational Longitudinal Study (NELS), Rowan et al. (1997) found that mathematics achievement was directly affected by a survey measure of teachers' mathematics knowledge, suggesting that teacher scores on subject matter tests may relate to student achievement as well. A few studies that examined pedagogical knowledge tests found that higher teacher scores were also related to higher student test performance, although many of these were dated (1979 or earlier). Strauss and Sawyer (1986) reported a modest and positive relationship between teachers' performance on the National Teacher Examination (NTE) and district average NTE scores, after controlling for size, wealth, racial/ethnic composition, and number of students interested in postsecondary education in the district.

The most recent literature on teacher quality uses panel data to better control for student heterogeneity and in some cases teacher heterogeneity. These modeling approaches link the current student achievement level to current family, teacher, and school inputs as well as to inputs provided in previous time periods. The methods examine the contribution of current education inputs (e.g., a school reform or assignment to a high "quality" teacher) on student outcomes. Ideally, model estimation requires a comprehensive history of all past and present family and school/teacher inputs as well as information about each student's endowed ability. This complete information is not available, however, leading to potential biases due to student unobserved heterogeneity, teacher unobserved heterogeneity and non random assignment of students to particular teachers (Todd and Wolpin, 2003). Two modeling approaches have been used in the literature to reduce these potential biases by incorporating unobserved prior year factors than may affect learning. In the contemporaneous value-added model, achievement scores are a function of current school and family inputs with student (and sometimes teacher) fixed effects to control for prior inputs. Alternatively, the value-added gains approach models *gains* in achievement scores as a function of current school and family inputs including also student (and sometimes teacher) fixed effects to control for prior inputs (Rivkin, 2006; Buddin and Zamarro, 2008).

Rivkin et al. (2005) is one of the earliest and perhaps most influential studies to estimate teacher effects from panel data (working drafts of the final report were available in 1998). The study uses longitudinal data on individual student achievement scores for Texas students in grades 3 through 6.[1] They use a value-added gains model with student and school fixed effects. Teacher quality has a large effect on student achievement in this study, but observed teacher qualifications like experience and education explain only a small share of the differences in teacher quality. In addition, they find that most of the variability in teacher quality was within

[1] The Texas data used in this analysis does not link students with individual teachers. The authors know the average characteristics of teachers by grade within each school and use these average teacher characteristics in their analysis.

schools and not across schools—an indication that high-performing teachers were not concentrated in a few schools.

Jacob and Lefgren (2008) examine how differences in teacher quality affected student achievement in a midsized school district. Like Rivkin et al. (2005), they find large differences in value-added measures of teacher effectiveness (teacher heterogeneity) but small effects of teacher qualifications like experience and education. They find that school principal rankings of teachers are better predictors of teacher performance than are observed teacher qualifications.

Harris and Sass (2006) examine how teacher qualifications and in-service training affected student achievement in Florida. They use a value-added gains model that controls for student and teacher fixed effects. They find small effects of experience and educational background on teacher performance. In addition, they find that a teacher's college major or scholastic aptitude (SAT or ACT score) is unrelated to their classroom performance.

Clotfelter et al. (2007) finds fairly similar parameter estimates for a variety of valued-added models for elementary students and teachers in North Carolina. They find that teacher experience, education, and licensure test scores have positive effects on student achievement. These effects are large (relative to socio-economic characteristics) for math, but the effects are smaller in reading.

Goldhaber (2007) also focuses on elementary students in North Carolina. He finds a small effect of teacher licensure test scores on student achievement. He estimates a value-added gains score model with lagged test score as a regressor. The author argues that raising the passing cut score would substantially reduce the pool of eligible teachers in North Carolina without having a substantial effect on student achievement scores.

Aaronson et al. (2008) looks at teacher quality and student achievement in Chicago public schools. Their study uses a gains score approach with controls for student and teacher fixed effects. The results show strong effects of teachers on student achievement, but traditional

measures of teacher qualifications like education, experience, and credential type have little effect on classroom results.

Koedel and Betts (2007) use a value-added gains model to look at student achievement of elementary students in San Diego. Like several of the other studies, they find that teacher quality is an important predictor of student achievement, but measured teacher qualifications (experience, quality of undergraduate college, education level, and college major) have little effect on student achievement.

The results from these studies are fairly consistent in showing that teacher quality has large effects on student achievement, but specific teacher qualifications have small effects on achievement (the exception is the one North Carolina study). Only the two studies with North Carolina data have information on teacher licensure scores. A concern for the results from these studies is the absence of controls for teacher heterogeneity. The assumption that schools or teachers are homogenous (no controlling for school unobserved heterogeneity or teacher unobserved heterogeneity) or that their differences can be controlled with observable characteristics has been contradicted by the evidence from the other studies. We argue that it is important to control for teacher heterogeneity to get consistent estimates of the student achievement model.

**Occupational Licensure**

Licensure tests are designed to improve the quality of the teaching workforce, but economic theory has offered two competing views of occupational licensure. Public interest theory suggests that regulation of occupation quality may be "needed" when quality is difficult to measure. Leland (1979) shows that salaries are set to average market quality of services when the quality of services is difficult to measure. High-quality providers switch to other occupations or reduce the quality of their services to match the quality of low-quality providers. While parents

may have poor information on teaching quality, district and school administrators should have more reliable estimates of teacher effectiveness.

Alternatively, "capture" theory suggest that occupations may use licensure regulation to limit occupational entry and inflate wages (Friedman and Kuznets, 1945; Schultz, 1982). Occupations generally design their own regulation and use this authority to limit entry into the profession. Wages rise because of the smaller pool of professionals in the occupation. Angrist and Guryan (2008) argue that licensure regulations may have an adverse effect on teacher quality, because regulation raises the costs for high-quality teaching applicants and may discourage them from entering the teaching profession.[2]

Our research does not test the merits of these theories. Rather, we focus on whether licensure scores predict teacher performance in the classroom. If these tests are improving the quality of the workforce, then licensure scores should be positively related to student test scores in the classroom. Two caveats are in order. First, the tests may set a floor for teaching quality and teachers below the exam cut scores might have much poorer success than teachers performing above the cut scores. Second, student test scores are an important measure of teaching success, but they are an imperfect measure of student learning.

## 3. ECONOMETRIC METHODS AND DATA

### Modeling Issues

We estimate both a contemporaneous value-added and value-added gains specifications that include student and teacher fixed effects in the following reduced forms:

$$Y_{it} = x_{it}\beta^C + u_i\eta^C + q_j\rho^C + \alpha^C_i + \phi^C_j + \varepsilon^C_{it} \quad \text{Contemporaneous Value-added}$$

$$Y_{it-} Y_{it-1} = x_{it}\beta^G + u_i\eta^G + q_j\rho^G + \alpha^G_i + \phi^G_j + \varepsilon^G_{it} \quad \text{Value-added Gains}$$

[2] On a similar note, Ballou and Podgursky (1995) argue that salary increases for teachers might have the perverse effect of reducing teaching quality. This could occur because higher wages reduce attrition from the current teaching workforce and reduce the number of entry level vacancies available.

where $Y_{it}$ is the test score (e.g. reading and math scores) of the student i in year t; $x_{it}$ are time-variant individual observable characteristics (classroom characteristics); $u_i$ are time-invariant individual observable characteristics (gender, race, parent's education, special attitudes and needs); $q_j$ are time-invariant observable characteristics of the jth teacher (gender, licensure test scores, education, experience); $\alpha^A_i$; A=C,G are individual time-invariant unobservables and $\phi^A_j$; A=C, G are teacher time-invariant unobservables. Finally, $\varepsilon^A_{it}$; A=C,G contains individual and teacher time variant unobserved characteristics.

Both teachers and students enter and exit the panel so, we have an unbalanced panel. Students also change teachers (generally from year to year). This is crucial, because fixed effects are indentified only for students who move. It is assumed that $\varepsilon_{it}$ is strictly exogenous. That is, student's assignments to teachers are independent of $\varepsilon_{it}$. Note, according to this assumption, assignment of students to teachers may be a function of the observables and the time-invariant unobservables.

It is usual to assume that the unobserved heterogeneity terms ($\alpha^A_i$; A=C,G and $\phi^A_j$; A=C, G) are correlated with the observables (due to student unobserved heterogeneity, teacher unobserved heterogeneity and non-random assignment of students to teachers). Thus, random effect methods are inconsistent and fixed effect methods are needed. In this case, the coefficients of students and teachers' time invariant observed characteristics ($\rho^A$ and $\eta^A$; $A=C,G$) are not identified separately from the unobserved heterogeneity terms. Given that the objective of this paper is to asses the role of such observed teacher characteristics on determining student performance, rather than dropping the variables $u_i$ and $q_j$, we define:

$$\psi^A_j = \phi^A_j + q_j\rho^A$$

$$\theta^A_i = \alpha^A_i + u_i\eta^A$$

Then, we estimate the models in two steps. In a first step we estimate the following equations using fixed effects methods:

$$Y_{it} = x_{it}\beta^{C} + \theta^{C}_{i} + \psi^{C}_{j} + \varepsilon^{C}_{it} \qquad \text{Contemporaneous Value-added} \quad (1)$$

$$Y_{it} - Y_{it-1} = x_{it}\beta^{G} + \theta^{G}_{i} + \psi^{G}_{j} + \varepsilon^{G}_{it} \qquad \text{Value-added Gains} \quad (2)$$

Then, in a second-stage regression we evaluate the ability of a rich set of observable teacher qualifications to predict teacher quality ($\psi^{A}_{j}$; $A=C,G$). Many of the observable teacher characteristics considered in this analysis are important determinants of teacher recruitment, retention and salaries decisions. For completion, in the same way, we also analyze the ability of observable student characteristics to predict student ability terms ($\theta^{A}_{i}$). Causal interpretation of the coefficients in these second step regressions would need the additional assumptions that $Cov(u_i,\alpha^{A}_{i})=Cov(q_j, \phi^{A}_{j})=0$. As explained above, this assumption is unlikely to be satisfied in this context. Thus, our second step estimates are measures of the correlation between observed characteristics and the teacher quality and student ability terms and not causal effects. Finally, our dependent variables in these second step regressions are statistical estimates of the true measures of teacher quality and student ability ($\psi^{A}_{j}$ and $\theta^{A}_{i}$) and as such they are measured with error. Thus, to obtain efficient estimates of the parameters we perform Feasible Generalized Least Squares (FGLS) regressions where the weights are computed following Borjas (1987).

A practical problem in estimating equations (1 and 2) is that there is no straight forward algebraic transformation of the observables that allow us estimate these equations and easily recover the estimates of the students and teachers' fixed effects.[3] Abowd et al. (1999), in an application for employer- employee data, propose to explicitly including dummy variables for employer heterogeneity and sweeping out the employee heterogeneity algebraically. They proved that this approach gives the same solution as the Least Squares Dummy Variables estimator for fixed effects panel data models. However, this method leads to computational difficulties because

[3] See Abowd et al (1999) for a description of suitable methods to estimate models with two levels fixed effects in the context of linked employer-employee data.

the software needs to invert a (K+J)×(K+J) matrix and store a lot of information. K refers to the total number of explanatory variables while J is the total number of teachers. Thus, we estimate the model using a preconditioned conjugate gradient method described in Abowd et al. (2002).[4]

Other potential data problems include, sample selection and attrition. Sample selection is due to the fact that we only observe teachers who passed their licensure exams. Although we acknowledge that the results we obtain are not representative for the whole population of potential teachers, they are for those teachers who are deemed eligible to teach. In this sense, we still believe the estimates we obtain in this population are the most relevant ones because these are the teachers who effectively will be participating in the educational system. On the other hand, literature suggests that more qualified teachers are more likely to leave the profession sooner (See e.g. Goldhaber (2007)). This phenomenon constitutes another source of potential bias. Following Goldhaber (2007) we also performed our estimates concentrating on a subsample of novice teachers. Results did not differ from the ones obtained for the whole sample, so we only present the results for the complete sample below.

**Data Issues**

*Student Achievement Data*

This study uses panel data from the Los Angeles Unified School District (LAUSD) for students in grades 2 through 5 for five consecutive school years from 2000 to 2004. The students are enrolled in self-contained classrooms taught by a single teacher, where the student and teacher data are linked by an identifying variable.[5]

This matched LAUSD student/teacher data are unusual in student achievement analysis. Districts often maintain separate administrative records for teachers and have difficulty linking

[4] The STATA routine used for this estimation was developed by Amenie Ouazad and is available on the web at http://repository.ciser.cornell.edu/viewcvs-public/cg2/branches/stata/.
[5] For privacy reasons, all teacher and student data in our analysis have scrambled identifiers. This allows the tracking of students and teachers overtime without compromising the privacy of individuals in the analysis.

students to individual teachers. Rivkin et al. (2005) are not able to match individual teachers with students and rely on the average characteristics of teachers in each grade and year for their study. Similarly, North Carolina data links students with the individual who proctored the test and not necessarily the student's teacher. Clotfelter et al. (2007) rely on an imputation strategy to link students with their classroom teacher. The authors were able to match about 75 percent of elementary math and reading teachers.

LAUSD is a large, diverse urban school district. Annual enrollment is about 730,000 students in over 800 schools.[6] Table 3.1 shows that 73 percent of elementary students are Hispanic, 11 percent are black, 10 percent are white/non-Hispanic, and 6 percent are Asian/Pacific Islander. Half of the students are classified as Limited English Proficient (LEP). About 80 percent of students are eligible for the free/reduced lunch program. While 33 percent of students have parents who did not graduate from high school, another 20 percent of students have a parent with a college or graduate school degree.

LEP students are predominantly Hispanic and from low-income families. The share of Hispanic, Asian/Pacific Islander, white non-Hispanic, and black students classified as LEP is 65, 31, 12, and 1 percent, respectively. About 57 percent of LEP students are eligible for free/reduced lunch as compared with 27 percent of non-LEP students. LEP students are much more likely to have parents without a high school diploma than other students.

California measures student achievement with the California Achievement Test, Sixth Edition (CAT/6) in reading and math. Test scores reflect student performance for each grade and subject relative to a representative national sample of students (norm group). Reading and math results are provided in a normal curve equivalent (NCE) scale, where the score ranges from 1 to 100 with a mean of 50. The average scores for LASUD students in our sample were 40 in reading and 47 in math.

[6] By way of comparison, LAUSD enrollment is larger than the enrollment in 28 states.

*Teacher Characteristics and California Licensure Test Data*

The elementary LAUSD teacher workforce is diverse and experienced. The average teaching tenure is 10 years, but the distribution is skewed with about 20 percent of teachers in their first three years of teaching. Three-fourths of the teachers are women. The race/ethnic distribution of teachers is 56 percent white non-Hispanic, 32 percent Hispanic, 12 percent black, and 12 percent Asian. About 20 percent of the teachers have a master's degree, but only 1 percent has a doctorate.

California requires new elementary teachers to pass up to three tests as part of state certification procedures (Le and Buddin, 2005).

- Basic Skills. Potential teaching candidates generally take the California Basic Educational Skills Test (CBEST) before admission to a teacher preparation program. The test focuses on proficiency in reading, writing, and mathematics.
- Subject-Matter Knowledge. Each candidate is required to show competence in the material in their authorized area. The state provides two types of subject matter tests using the California Subject Examinations for Teachers (CSET): a multiple subject exam for elementary school teachers and a single subject exam for middle and secondary school teachers. These skills are acquired in subject-matter departments and outside of teacher preparation programs.[7]
- Reading Pedagogy. The Reading Instruction Competence Assessment (RICA) is required for all elementary school teachers. This is the only licensure test that specifically assesses skills acquired through professional teacher preparation programs.

Over 80 percent of white, non-Hispanic and Asian/Pacific Islander teaching candidates in California pass each test on the first attempt, but far fewer Black and Hispanic do so (Jacobson and Suckow, 2006). The pass rates for Hispanics are 53, 60, and 72 percent in basic proficiency,

[7] Prior to NCLB legislation in 2001, teaching candidates could demonstrate subject-matter knowledge by either passing the state mandated licensure test or by completing an approved subject matter preparation program. Under NCLB, candidates are required to pass a subject matter test.

subject area knowledge, and reading pedagogy respectively. For black/African American candidates, the first-time pass rates are 44, 48, and 67 in basic proficiency, subject matter knowledge, respectively.

After retesting, the pass rates increase substantially, and the race/ethnic gap in pass rates narrows considerably. This suggests that many candidates may improve their skills and preparation to meet the pass criterion or test familiarity boosts scores. The cumulative pass rates for white non-Hispanics are 93, 87, and 97 percent in basic proficiency, subject area knowledge, and reading pedagogy, respectively. The corresponding rates for blacks are 69, 65, and 88 percent, and the rates for Hispanics are 77, 72, and 92 percent. Many candidates may be discouraged by failing one of the tests, however, and lose interest in teaching.

The California Commission on Teacher Credentialing collects licensure test scores as part of teacher certification procedures. Individuals are informed of their passing status on test scores and subtests. Districts are not informed of licensure test scores, but they are informed when a teacher completes certification requirements for a multiple-subject credential (elementary school teachers) or single-subject credential (middle- and high-school teacher).

We worked with the California State University (CSU), Chancellor's Office, to obtain teacher licensure scores for six cohorts of teachers from the CSU system (years 2000 through 2006). The file includes licensure scores for about 62,000 teaching candidates. Separate scores are recorded on a basic skills test, subject area tests, and reading pedagogy. The file contains information on failed exams, so we know whether a teacher needed to retake one or more exams as part of the certification process.

The CSU licensure data are available for 17 percent of LAUSD teachers in our analysis sample (2738 matches of 16,412 teachers). This low match rate reflects two key factors. First, most teachers in the district received their certification before 2000 and have been teaching for some time. The match rate rises to 38 percent for teachers in their first three years of teaching. Second, CSU only has access for licensure scores for candidates from their various campuses and

not from the entire state. About 50 percent of California teaching certificate completers are affiliated with a CSU campus. We were unable to obtain additional licensure information from either the California Commission on Teacher Credentialing or other campuses.

We used several different empirical methods to handle the missing information on licensure test scores. In each approach, we estimated stage 1 regressions as described above on the entire sample. The adjustment for missing licensure data occurs in stage 2 using data on estimated teacher effects in reading and math.

- Multiple imputations. This approach imputes licensure scores from other teacher characteristics and estimated teacher effects in reading and math. Different datasets are created with different imputed values, and final parameters estimates are blended from regressions on each dataset. The methods rely on assumptions such as Missing at Random or Missing Completely at Random that are made on the conditional distributions of the licensure score variables.[8] We are concerned that this approach is not well suited to our situations where we have large proportions of missing variables, and we would rather prefer not to make assumptions about their (conditional) distributions.
- Dropping records with missing teacher data. In this approach, we estimate stage 2 entirely on matched CSU teachers. The results show whether licensure scores for recent CSU teaching graduates are significantly related to student achievement in each teacher's classroom. This approach focuses on the CSU sample of young teachers and ignores the other teachers. The broader group of teachers would provide more information on how other teacher characteristics affect student achievement.
- Missing dummy variables. A common missing value adjustment consists of setting the value of the missing covariate to an arbitrary fixed value (zero) and, adding dummy variables for "missings."

[8] See, e.g., Rubin (1996) for a description of Missing at Random and Missing Completely at Random assumptions and their application in imputing methods.

The main analysis results reported below rely on dropping records with missing teacher data. We found similar results with the other methods, so the teacher licensure results were robust across the alternative methods of handling the missing values.

*Patterns of Student and Teacher Characteristics across Schools*

Test scores vary considerably across different types of students and different schools in LAUSD. Table 3.2 shows the simple patterns in student and teacher characteristics for schools in the lowest test score quartile as compared with the highest test score quartile. The test score gap is 20 percentage points in reading and 22 points in math. These differences may reflect differences in the background and preparation of students attending different schools as well as the quality of instruction at each group of schools. Low-scoring schools have much higher concentrations of black, Hispanic, and LEP students than do higher scoring schools. In addition, family socioeconomic status is much lower in the lowest quartile schools, where nearly 50 percent of students have parents without a high school degree.

Teacher characteristics also vary considerably with average school test score, reflecting some sorting of teachers into schools. Low-scoring schools have more new teachers and a less experienced teacher workforce than high-scoring schools. Fewer teachers in low scoring schools have advanced degrees, perhaps reflecting the low experience mix in these schools. Black and Hispanic teachers are much more common in low-scoring schools. Finally, teacher licensure scores are consistently lower in the lowest quartile schools relative to the highest quartile schools.

The teacher assignment patterns hint that differences in student achievement might reflect the assignment of lower quality teachers to schools with more at-risk students. The patterns show that the schools with the most at-risk students have more new teachers, fewer teachers with advanced degrees, and teachers with lower teacher licensure test scores. The next section will begin to disentangle how these teacher characteristics translate into student achievement outcomes.

## 4. RESULTS

This section presents the results from the values-added models of student achievement. We divided the results into four subsections. The first examines the distribution of student and teacher quality across schools in the district. The second subsection shows the results of the stage 1 regressions for time-varying variables. Subsections three and four examine factors affecting teacher and student heterogeneity, respectively.

### Teacher quality and school quality contributions to student performance

Are "good" teachers concentrated in a few schools (presumably with few at-risk students), or are high-quality teachers distributed broadly across a variety of schools? Table 4.1 shows the results of fixed effects regressions for unconditional models that adjust only for grade and test year. The results show that student-to-student deviations in achievement are about four times as large as teacher-to-teacher deviations.[9] A typical student assigned to a teacher one standard deviation above the mean is expected to score about 5 or 6 percentage points higher in reading and math, respectively, than a comparable student assigned to an average teacher (the teacher effect size is about 0.2).

School effects are much smaller than teacher effects. The second model in Table 4.1 shows a baseline model that controls for student and school effects. The results show achievement for comparable students differs much less from school to school than it does from teacher to teacher in the first model. A standard deviation school "quality" is associated with about 2 percentage point differences in student achievement (a school effect size of about 0.1).

The results from Table 4.1 indicate that high-quality teachers are not concentrated in a few schools. School effects are much smaller than teacher effects, and this indicates the

[9] Standard errors of student, teacher and school fixed effects presented in this table correct for the sampling error due to the fact that these terms are estimates. Jacob and Lefgren (2008) provide a detailed description of the empirical Bayes procedure used to eliminate attenuation bias.

dispersion of high-quality teachers (as measured by their effects on individual student achievement) across schools. This dispersion collapses much of the variance in outcomes at the school level, because individual schools are composed of a mix of low- and high-quality teachers.

These simple models provide a broad description of how student achievement varies across students and teachers. We now turn to models that decompose linkages between student achievement outcomes and individual student and teacher factors.

**Estimates of Value-Added Models**

Table 4.2 shows the results for the contemporaneous value-added model (levels) and the value-added gains model (gains). Each model version controls for test year and grade as well as for time-varying student and classroom characteristics. In addition, each specification includes student and teacher fixed effects. The time-varying factors consist of three types of components: class size, class peer composition, and student/teacher match variables. Peer effects measures are the proportion of different ethnicity groups and female students in the classroom. As explained in previous sections, the central problem with estimating the effect of these peer and match variables is that families may self-select their children into classrooms and schools depending on their children ability. Moreover, schools may assign their teachers to a given classroom depending on its composition. As a result, these variables are potentially endogenous due to this non random assignment of students into classrooms. Our estimates use both student and teacher fixed effects to allow for correlation between them and the explanatory variables.[10]

[10] Most of the research on peer effects dealt with selection by controlling for observable variables, comparing siblings that experienced different schools, examining desegregation programs or estimating selection models (Angrist & Lang, 2002). Other parts of the literature exploit the availability of policy or natural experiments to estimate peer effects (Zimmerman, 1999 and Sacerdote, 2000). Hoxby (2000) exploits the variation in adjacent cohorts' peer composition within a grade within a school that is idiosyncratic to estimate peer effects. Cullen and Jacob (2007) use lottery data to look at open enrollment effects for Chicago elementary school students. They find lottery winners are matched with higher quality peers in their new schools but their subsequent achievement scores are not higher than those of lottery losers.

The results between reading and math are similar in both models, but more factors are more significant in the levels model than in the gains model. Class size has a negative and significant effect in all specifications for both reading and math scores. The magnitude of the effect is small, however, since a five-student drop in class size would only increase reading and math levels by about one percentage point. Nearly all of the peer effect and student/teacher match variables are insignificant in the gains model. Gain scores are significantly lower in math for classes with a larger share of black students. In the levels, model, the proportion of girls has a positive effect on achievement in both reading and math. The proportion of blacks is inversely related to both reading and math scores. The proportion of Hispanics is inversely related to achievement in reading (perhaps reflecting language difficulties), but the effect is not significant in math.

The results provide little evidence that students have higher achievement levels if matched with a similar teacher. Dee (2005), Clotfelter et al. (2007), and Ouazad (2007) find that students do better academically when they are matched with a teacher of similar race/ethnicity or gender. None of the student/teacher match variables are significantly different from zero in the gains specification in Table 4.2, and few match variables are significant in the levels model. Black students have higher math scores if matched with a black teacher, but all other race/ethnicity matches are insignificant. Female students have lower reading and math scores in levels when matched with a female teacher.

Table 4.3 describes details of the distribution of empirical Bayes estimates of teacher fixed effects. The range of teacher effects is large—the interquartile range (the $25^{th}$ to $75^{th}$ percentile) is about 5 to 7 points in levels and 8 to 12 points in gains. The skewness measures indicate that in all cases but in the case of reading scores for the levels specification the distribution of teacher fixed effects has slightly more mass probability in the left of the distribution than a normal distributed variable (skewness=0). On the other hand, the kurtosis

coefficients indicate that the distributions of teacher fixed effects have, in all cases, higher probability than a normally distributed variable of values near the mean.

**Teacher Quality and Observed Teacher Characteristics**

Second-stage regressions show how time-invariant teacher characteristics affect student achievement in the classroom. Teacher characteristics include experience, gender, race/ethnicity, education level, and teacher licensure test scores.

As we can see in Table 4.4 licensure test results for different tests are highly correlated, especially for CSET and CBEST results. To avoid problems of multicollinearity and to provide a clearer interpretation of the results, we estimated linear regression models for each of the licensure test results both jointly and separately.

Tables 4.5 and 4.6 show the results for reading and math student tests results obtained for the levels specification. Teacher experience has a positive effect on student achievement in each specification for reading and math, but the effect is small. A five-year increase in teacher experience is associated with only a 0.5 and a 0.8 percentage point increase in reading and math scores, respectively. Female teachers have better student outcomes than males—comparable students score about one percentage point higher in reading and math with female teachers than with male teachers. Teachers with masters or a doctorate degree do no better or worse in either reading or math than comparable teachers without advanced degrees.

Teacher race/ethnicity has a stronger effect on math achievement than on reading achievement. Students with an Asian/Pacific Islander teacher do better in reading than with a white non-Hispanic teacher. Black and Hispanic reading teachers are not significantly different than white non-Hispanic teachers. In math, the differences are larger. Black math teachers have classroom scores about 0.7 percentage points lower than white non-Hispanic teachers. Hispanic and Asian/Pacific Islander math teachers have scores 0.4 and 1.3 percentage points higher than non-Hispanic teachers.

The teacher licensure scores have little if any effect on classroom student achievement. CBEST, CSET, and RICA are all insignificant in the reading models. In math, CBEST and CSET are significant and negative, i.e., better licensure scores are associated with *lower* student achievement scores in the classroom. In both cases, the effect is small, however, with a one standard deviation change in test scores linked with a half point reduction in classroom achievement. RICA does have a small positive effect on student achievement in math, but this effect is only significant in the model with all three licensure tests combined.

Tables 4.7 and 4.8 show the gains model regressions for teacher heterogeneity. The pattern of results is similar to that for levels. Classroom achievement is increasing at a decreasing rate with teacher experience in both reading and math. Female teachers have better classroom results than male teachers. Advance educational degrees are not associated with better classroom achievement results. Teacher race/ethnicity has similar effects in both reading and math. Black teachers have lower classroom results than white non-Hispanics, while Hispanic and Asian/Pacific Islander teachers have better results than white non-Hispanics.

Teacher licensure tests have little effect on student achievement. None of the licensure scores are significant in the joint reading specification. The licensure test is only significant in the separate CSET model, and there it has a negative sign (more subject matter knowledge is associated with slightly lower student achievement in reading). In the math models of Table 4.8, the licensure scores are all insignificantly different from zero or have the wrong sign. The separate specifications of CBEST and CSET show that test scores vary inversely with math student achievement in the classroom.

Linear regression models were also estimated restricting the sample to novice teachers and classifying teachers according to the grade they have been teaching most frequently, allowing for different effects of licensure test results depending on different grades. Results did not differ substantially from the ones we just presented. In addition, we estimated the levels and gains models with indicator variables for whether the teacher has initially failed either CBEST, CSET,

or RICA tests (see Appendix Tables A-1 and A-2). The results indicated that teachers who failed one or more of the licensure tests did neither better or worse in the classroom than did teachers who passed the exams on the first try. In addition, we tested for non-linearities in the licensure scores and found these variables were also insignificantly different from zero (see Appendix Table A-3).

The teacher results were robust across a broad range of specifications. Teacher experience had a weak positive effect on student achievement that declines over time. Teacher gender and race/ethnicity have some effects on achievement. Advanced teacher educational degrees have no bearing on student achievement. Student achievement scores are not significantly affected by the basic skills, subject matter, or reading pedagogy skills of their teachers as measured on current California licensure tests.

The estimated teacher effects have important implications for the patterns of sorting shown in Table 3.2. Teachers with lower experience, education, and licensure test scores are concentrated in schools with poor average test score, but our results show that these differences contribute little to a school's student achievement. In addition, the distribution of unobserved teacher heterogeneity across schools is much more balanced across schools than is the distribution of measured teacher attributes. The gap between average teacher heterogeneity of the lowest and highest quartile schools is only 0.9 and 1.3 percentage points in reading and math, respectively. This finding means that, on average, teachers are making comparable improvements across a broad range of schools and that different performance in these schools is mostly due to student characteristics.

*Comparison to Models with Only Student Fixed Effects*

Clotfelter et al. (2007) and Goldhaber (2007) estimated the effects of teacher licensure scores on student achievement using North Carolina data. These studies find positive effects of North Carolina licensure tests, but these models control for student heterogeneity and not teacher

heterogeneity. These types of models implicitly assume that schools or teachers are homogenous or that they differ in a few observable dimensions. Rivkin et al. (2005) show weak linkages between observable teacher characteristics and classroom performance. Similarly, the results in Tables 4.5 through 4.7 show that observable teacher characteristics like those used by Clotfelter et al. (2007) and Goldhaber (2007) explain only a small portion of teacher heterogeneity across classrooms. As a result, not controlling for this source of unobserved heterogeneity can lead to important sources of biases.

For comparison purposes, we reestimated the joint models from Tables 4.5 through 4.8 under the assumption of student heterogeneity and not teacher heterogeneity. The comparisons in Table 4.9 indicate that estimated teacher parameters are sensitive to the inclusion of teacher heterogeneity controls. Experience effects are much weaker in all specification in Table 4.9 than in the corresponding models with teacher heterogeneity controls. Similarly, the gender and race/ethnicity effects are smaller in the restricted models than in the more general ones. Teacher education remains insignificant in each equation of Table 4.9. Among licensure test scores, RICA is significant and positive in the levels models with only student controls, but the effect of RICA is insignificant in the gains models. CBEST is significant and negative in both models in Table 4.9.

The results in Table 4.9 show that the inclusion of teacher heterogeneity is important for estimating the contributions of teachers to student learning, but the results also highlight underlying differences between the North Carolina and California data. With or without teacher heterogeneity, our measures of observed teacher characteristics are weaker than those reported by Clotfelter et al. (2007). Similarly, our estimates of teacher licensure effects on student achievement are weaker than those of either Clotfelter et al (2007) or Goldhaber (2007).

**Student Quality and Observed Student Characteristics**

Table 4.10 shows how observed student characteristics explain differences in student unobserved heterogeneity. The explanatory variables are gender, race/ethnicity, LEP indicator, whether the student receives free/reduced school lunch, parent's education, controls, and indicators for student enrollment in a gifted or special education program. The table includes reading and math specifications for both the levels and gains models.

The levels results show large differences in achievement scores across different student types. In general, black and Hispanic students have lower scores than non-Hispanic white students. Asian/Pacific Islander students have lower reading and higher math scores than do non-Hispanic whites. LEP students do consistently worse than others, but the gap is smaller in math than in reading (presumably reflecting the lower language emphasis in math). Girls do better in reading and worse in math than do boys.

Socioeconomic status is a strong predictor of student success. Students in the free/reduced lunch program have lower scores in both reading and math than do similar others. Reading and math scores are positively related to education—the lowest performing students have no parent with a high school diploma and student scores rise consistently with each increment in parental education. Greater family wealth may affect students through greater resources in the home to complement schoolhouse learning. Alternatively, these parents may place greater emphasis on student learning or provide more support for their children. Finally, gifted and special education students have much different scores than other students. These variables are included as controls and have the expected effects.

The gains results generally mirror those of the levels model, but fewer factors are statistically significant. One issue for the gains model is that little student-level heterogeneity remains after computing the gains score and remaining student effects reflect differences in growth rates for particular groups. The results show that black students have lower growth than

white non-Hispanics. Hispanic students have lower growth in math (but not in reading) than white non-Hispanics. LEP students have higher growth in reading (but not in math) than English proficient students—this may reflect students "catching up" as they become more proficient in English. Girls have higher growth rates than boys in both reading and in math.

Socioeconomic status has less effect on growth than on levels. Free/reduced lunch students have higher growth in reading than others, but the growth effect is insignificant in math. Growth rates are positively related to parental education, but the effect is only significant for families where the parents have a college or graduate school degree.

Finally, growth rates seem to be lower for gifted students and higher for special education students. The reasons for these effects are unclear. Perhaps gifted students enter the program after a very strong year and then regress to the mean. Special education students may be improving and learning to adapt to their problems. The gifted and special education programs are not a focus of this study, and further investigation is needed to sort out how and why these students have these achievement patterns.

## CONCLUSIONS AND IMPLICATIONS

Teacher quality is an important determinant of student achievement, but measured teacher qualifications and preparation explain little of the observed differences in student outcomes across teachers. This poses a dilemma for educators and policy makers—while teachers have large effects on student achievement, the research evidence provides little indication how teacher quality can be enhanced. Future research should focus on the identification of specific teacher attributes or practices that enhance with student achievement in the classroom.

*How does teacher quality vary across classrooms and across schools?*

The evidence shows considerable variation in teacher quality across classrooms and less variation across schools. A one-standard-deviation change in teacher quality is associated with a 5 or 6 percentage point increase in reading and math achievement (an effect size of 0.2). The variation in school quality is about half as large—an indication that low- and high-quality teachers are not separated into two disjointed sets of schools.

Traditional measures of teacher quality vary substantially between schools with low- and high-test scores. Schools in the lowest student achievement quartile have teachers with lower average experience, lower educational preparation, and lower scores on teacher licensure tests than do schools in the highest student achievement quartile. However, these traditional teacher quality measures explain little of the learning gap between these schools. Rather, the teacher quality gap between low- and high- scoring schools is only about one percentage point.

Implications of teacher quality distribution across schools

These results have important implications for improving test scores in low performing schools. Efforts to improve the teaching performance in these schools are unlikely to succeed, if they rely entirely on traditional measures of teacher quality (Steel, 2007). A simple reshuffling of teachers is unlikely to produce substantial achievement improvement in low-performing schools.

*Do traditional measures of teacher quality like experience and teacher educational preparation explain their classroom results*?

Teacher experience is weakly related to student achievement, and teacher education level has no effect on student achievement. These results are consistent across a variety of specifications of the contemporaneous value-added and gains value-added models.

High levels of teacher experience may have important benefits for schools, even if teacher experience is weakly related to student achievement. Longer teacher retention saves

money in recruiting and training teachers. These savings may indirectly affect resources that are ultimately available for classroom instruction and improved student achievement.

Similarly, advanced teacher degrees may have indirect benefits for the teacher workforce. Ongoing training may infuse a knowledge base of new teaching techniques that spill over to fellow teachers who are not enrolled in degree programs.

Implications of measures of teacher quality results

The current pay structure for teachers in the U.S. is input-based—teachers are paid on the basis of their skills, which are measured by education and teaching experience (Lazear 1986, 2000). The premise is that these input measures are ultimately linked to desired outcomes (i.e., more student learning or skills). But the evidence shows that these traditional input measures are weakly linked with student achievement.

Merit pay systems might realign teaching incentives by directly linking teacher pay with classroom performance (Buddin et al., 2007). Merit pay is "results oriented" in the sense that compensation focuses on the production of specific student outcomes. The challenge for designing a merit pay system for teachers is in defining an appropriate composite of student learning (output) and in measuring teacher performance in producing learning.

Ideally, merit pay would improve the teacher workforce in two ways. First, teachers would have incentives to increase effort to produce specific student outcomes. Second, linking pay directly with classroom outcomes would encourage high-quality teachers to remain in the teaching.

Output-based pay may also have adverse incentives on the teacher workforce. First, some tasks inherently involve team production, so individual contributions are difficult to disentangle. A compensation system could reward team output, but this would create incentives for individuals on the team to "free ride" on the efforts of others. Second, individual rewards for quantity produced will encourage undue emphasis on quantity alone in some circumstances. For example, if teachers received bonuses for the number of students reaching a reading proficiency

level, then they would have little incentive to focus on student above the proficiency level. Similarly, teachers might simply "teach the test" at the expense of promoting long-term learning. Third, most employees like predictable earnings and dislike large fluctuations in income. This suggests that merit pay for teachers should comprise a portion of teacher pay, but not the bulk of teacher compensation.

*Does teacher success on licensure test exams translate into better student achievement in a teacher's classroom?*

The results show no indication that any of the teacher licensure scores affect student achievement. The measured basic skills, subject-matter knowledge, and reading pedagogy scores of elementary teachers are unrelated to student achievement.

A limitation of the data is that licensure tests and teacher performance are available only for teachers who pass the tests. Licensure tests are designed to set minimum teaching proficiency standards. Potential teachers who fall below the cut scores on the licensure tests might indeed have worse classroom outcomes than teachers who ultimately surpass those cut scores.

Implications of licensure exam results

Different test content might change the measured relationship with student achievement. Le and Buddin (2005) show that the content of the California licensure examinations is appropriate for the teaching profession and the cut scores are established in a defensible manner. Other states use similar approaches to developing licensure exams, but they may be more successful than California in differentiating factors that affect classroom performance. The California evidence suggests that education experts should rethink the knowledge requirements for new teachers and develop tests that more accurately predict classroom performance. Different standards might restrict entry into the teacher profession, however, and have adverse consequences for teacher supply (Angrist and Guryan, 2008).

An alternative explanation for the weak effects of teacher quality measures on student achievement is that teaching effort is inversely related to those quality measures. More experienced or better educated or more skilled teachers (as measured by licensure exams) may inherently be better *able* to teach, but they may not persistently practice those abilities in the classroom. The current compensation system rewards measured teacher inputs and not performance per se. Perhaps this system provides too little incentive for the "best" teachers to deliver their best performance in the classroom on a consistent basis.

**REFERENCES**


Aaronson, D., Barrow, L., Sander, W., 2007. Teachers and Student Achievement in the Chicago Public High Schools, Journal of Labor Economics, 25(1), 95-135.

Abowd, J., Kramarz, F., Margolis, D., 1999. High Wage Workers and High Wage Firms, Econometrica, 67 (2), 251-333.

Abowd, J., Creecy, R., Kramarz, F., 2002. Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data, working paper.

Angrist, J., Lang, K., 2002. How Important Are Classroom Peer Effects? Evidence from Boston's Metco Program, NBER working Paper 9263.

Angrist, J., Guryan, J., 2008. Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements, Economics of Education Review, 27, 483-503.

Ballou, D., Podgursky, M., 1995. Recruiting Smarter Teachers, Journal of Human Resources, 30(2), 326-338

Borjas, G., 1987. Self-Selection and the Earning of Immigrants, American Economic Review, 77(4), 531-553.

Boyd, D., Lankford, H., Loeb, S., Wyckoff, J., 2005. The Draw of Home: How Teachers' Preferences for Proximit Disadvantage Urban Schools, Journal of Policy Analysis and Management, 24(1), 113-132.

Buddin, R., Zamarro, G., 2008. Teacher Quality, Teacher Licensure Tests, and Student Achievement, Working Paper, RAND WR-555-IES.

Buddin, R., McCaffrey, D., Kirby, S., Xia, N., 2007. Merit Pay for Florida Teachers: Design and Implementation Issues, Working paper, RAND WR-508-FEA.

Clotfelter, C., Ladd, H., Vigdor, J., 2007. How and Why Do Teacher Credentials Matter for Student Achievement? NBER Working Paper 12828.

Cullen, J., Jacob, B., 2007. Is Gaining Access to Selective Elementary Schools Gaining Ground? Evidence from Randomized Lotteries, NBER Working Paper 13443.

Dee, T., 2005. A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? American Economic Review, 95(2), 558-565

Ehrenberg, R., Brewer, D., 1995. Did teachers' verbal ability and race matter in the 1960's? Coleman revisited, Economics of Education Review, 141, 1-21.

Ferguson, R., Ladd, H., 1996. How and why money matters: An analysis of Alabama schools, In H.Ladd (Ed.), Holding schools accountable: Performance-based reform in education, 265-298. Washington, DC: Brookings Institute.

Ferguson, R., 1991. Paying for public education: New evidence on how and why money matters, Harvard Journal on Legislation, 282, 465-498.

Friedman, M., Kuznets, S., 1945. Income from Independent Professional Practice, NBER.

Goldhaber, D., 2007. Everyone's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness? Journal of Human Resources, 42(4), 765-794.

Harris, D., Sass, T., 2006. The Effects of Teacher Training on Teacher Value-Added, Working paper, Florida State University.

Hanushek, E., 1986. The Economics of Schooling: Production and Efficiency in Public Schools, Journal of Economic Literature, 24(4), 1141-1177.

Hoxby, C., 2000. Peer effects in the classroom: Learning from gender and race variation. Working Paper 7867, NBER.

Jacob, B., Lefgren, L., 2008. Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education, Journal of Labor Economics, 26(1), 101-136.

Jacobson, P., Suckow, M., 2006. Report on Passing Rates of Commission-Approved Exams For 2000-01 to 2004-05, California Commission on Teacher Credentialing.

Koedel, C., Betts, J., 2007. Re-Examining the Role of Teacher Quality in the Educational Production Function, Working paper, University of California, San Diego.

Lazear, E., 1986. Salaries and Piece Rates, Journal of Business, 59(3), 405-31.

Lazear, E., 2000. Performance Pay and Productivity, American Economic Review, 93(5), 1346-61.

Le, V., Buddin, R., 2005., Examining the Validity Evidence for California Teacher Licensure Exams, Working paper, RAND WR-334-EDU.

Leland, H., 1979. Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards, Journal of Political Economy, 87(6), 1328-1346.

Murnane, R., Steele, J., 2008. What Is the Problem? The Challenge of Providing Effective Teachers for All Children, Project Muse, 17(1), 15-43.

Ouazad, A., 2007. Assessed by a Teacher Like Me: Race, Gender and Subjective Evaluations, Working paper, London School of Economics.

Rivkin, S., 2006. Cumulative Nature of Learning and Specification Bias in Education Research, unpublished paper.

Rivkin, S., Hanushek, E., Kain, J., 2005. Teachers, Schools, and Academic Achievement, Econometrica, 73(2), 417-458.

Rowan, B., Chaing, F., Miller, R.J., 1997. Using research on employees' performance to study the effects of teachers on students' achievement, Sociology of Education, 704, 256-284.

Rowan, B., Correnti R., Miller, R. J., 2002. What large-scale research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools, Teachers College Record, 104(8), 1525-1567.

Rubin, D., 1996. Multiple Imputation after 18+ Years, Journal of the American Statistical Association, 91, 473-489.

Sacerdote, B., 2001. Peer effects with random assignment: Results for Dartmouth roommates, Quarterly Journal of Economics, 116(2), 681–704.

Schultz, T., 1982. Investing in People: The Economics of Population Quality, University of California, Berkeley, CA.

Steele, J., 2008. Do Financial Incentives Help Low-Performing Schools Attract and Keep Academically Talented Teachers? Evidence from California, Working paper, Harvard University.

Strauss, R., Sawyer, E., 1986. Some New Evidence on Teacher and Student Competencies, Economics of Education Review, 5(1), 41-48.

Todd, P., Wolpin, K., 2003. On the Specification and Estimation of the Production Function for Cognitive Achievement, Economic Journal, 113, F3-F33.

Wright, S., Horn, S., Sanders, W., 1997. Teacher and classroom context effects on student achievement: Implications for teacher evaluation, Journal of Personnel Evaluation in Education, 11, 57–67.

Zimmerman, D., 1999. Peer effects in academic outcomes: Evidence from a natural experiment. Working paper, Williams College.

**Table 3.1—Characteristics of Elementary Students**

| Student Characteristics | Proportion | LEP Share |
|---|---|---|
| White Non-Hispanic | 0.09 | 0.12 |
| Black | 0.11 | 0.01 |
| Hispanic | 0.73 | 0.65 |
| Asian/Pacific Islander | 0.06 | 0.31 |
| Female | 0.50 | 0.50 |
| Limited English Proficiency | 0.51 | ---- |
| Free/reduced lunch | 0.79 | 0.57 |
| *Highest Parental Education* | | |
| Not high school graduate | 0.33 | 0.72 |
| High school diploma | 0.28 | 0.51 |
| Some college | 0.17 | 0.32 |
| College graduate | 0.14 | 0.25 |
| Some graduate school | 0.06 | 0.23 |

**Table 3.2--Comparison of Student and Teacher Characteristics in Schools with Lowest and Highest Test Scores in 2004**

| School Characteristic | Lowest Quartile Schools | Highest Quartile Schools |
|---|---|---|
| Reading Percentile | 34.10 | 53.66 |
| Math Percentile | 40.79 | 62.31 |
| *Student Characteristics* | | |
| Black | 0.15 | 0.10 |
| Hispanic | 0.83 | 0.36 |
| LEP | 0.64 | 0.20 |
| Parents not high school graduates | 0.47 | 0.11 |
| *Teacher Characteristics* | | |
| Years of Experience | 6.36 | 9.37 |
| Experience < 3 yrs | 0.44 | 0.30 |
| Black | 0.21 | 0.08 |
| Hispanic | 0.37 | 0.14 |
| Master's/Doctorate | 0.16 | 0.23 |
| CBEST (standardized) | -0.52 | -0.08 |
| CSET (standardized) | -0.43 | 0.06 |
| RICA (standardized) | -0.31 | -0.01 |

Note: All factors differ significantly between the two groups of schools.

**Table 4.1—Comparison of Student, Teacher, and School Fixed Effects**

| | Reading | Math |
|---|---|---|
| *#1. Student & Teacher Fixed Effects* | | |
| Student ($\sigma_{Student}$) | 16.75 | 18.33 |
| Teacher ($\sigma_{Teacher}$) | 4.99 | 6.25 |
| *#2. Student & School Fixed Effects* | | |
| Student ($\sigma_{Student}$) | 16.97 | 18.69 |
| School ($\sigma_{School}$) | 2.15 | 2.57 |

**Table 4.2—Estimates of Contemporaneous Value-Added and Value-Added Gains Models**

| | Levels | | Gains | |
|---|---|---|---|---|
| Variable | Reading | Math | Reading | Math |
| Test Year 2001 | 4.7992* | 4.7409* | NA | NA |
| | (0.0539) | (0.0621) | | |
| Test Year 2002 | 8.7472* | 10.1358* | -1.82* | 0.6902* |
| | (0.0813) | (0.0999) | (0.118) | (0.1139) |
| Test Year 2003 | 8.8283* | 11.2429* | -5.7058* | -3.6568* |
| | (0.1221) | (0.1406) | (0.2197) | (0.2042) |
| Test Year 2004 | 11.4256* | 14.5627* | -0.3141 | -0.5286 |
| | (0.1454) | (0.1647) | (0.3033) | (0.2965) |
| Class Size | -0.1677* | -0.2224* | -0.0795* | -0.1306* |
| | (0.0065) | (0.0059) | (0.0148) | (0.0157) |
| Percent Female in Class | 0.4042* | 1.0647* | 0.248 | 1.2103 |
| | (0.2029) | (0.2117) | (0.4413) | (0.6601) |
| Percent Black in Class | -1.3819* | -1.8051* | -0.5991 | -2.3175* |
| | (0.4378) | (0.4616) | (1.0337) | (1.0983) |
| Percent Hispanic in Class | -0.9909* | -0.1097 | -1.2005 | 0.5385 |
| | (0.3318) | (0.3819) | (0.973) | (0.9165) |
| Percent Asian/Pacific Islander in Class | 0.0988 | -0.0768 | -1.3636 | -0.5706 |
| | (0.4465) | (0.5338) | (1.2689) | (1.239) |
| Hispanic Student & Teacher | -0.0755 | 0.0856 | -0.066 | 0.1476 |
| | (0.1322) | (0.1332) | (0.284) | (0.2923) |
| Black Student & Teacher | 0.1833 | 0.2393* | 0.5294 | 0.3505 |
| | (0.1327) | (0.1169) | (0.3705) | (0.3631) |
| Asian/Pacific Islander Student & Teacher | -0.1925 | -0.0576 | -0.677 | 0.0635 |
| | (0.1538) | (0.1918) | (0.3707) | (0.3737) |
| Female Student & Teacher | -0.1982* | -0.3269* | -0.0445 | 0.0176 |
| | (0.0614) | (0.0556) | (0.0982) | (0.1474) |
| College Parents & Teacher Masters/Ph.D. | 0.0242 | 0.0029 | 0.0286 | 0.0576 |
| | (0.0736) | (0.0878) | (0.2207) | (0.2213) |
| Standard Deviation of Student Effects | 17.08 | 18.82 | 8.98 | 10.32 |
| Standard Deviation of Teacher Effects | 5.07 | 6.65 | 11.04 | 14.02 |
| Number of Observations | 935,775 | 935,775 | 585,325 | 585,325 |
| Number of Students | 332,538 | 332,538 | 325,521 | 325,521 |
| Number of Teachers | 16,412 | 16,412 | 13,047 | 13,047 |

Note: Bootstrapped Standard errors are in parenthesis. An asterisk indicates significance at a 95% level. Controls for grades are also included.

**Table 4.3—Distributions of teacher effects**

| | Levels | | Gains | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| Mean | 0.04 | -0.12 | 2.19 | 1.25 |
| S.D | 4.67 | 6.16 | 9.52 | 12.47 |
| Skewness | -0.074 | 0.68 | 0.64 | 0.90 |
| Kurtosis | 7.25 | 4.52 | 12.84 | 9.30 |
| | | | | |
| Percentile | | | | |
| 5% | -6.73 | -9.07 | -10.09 | -15.32 |
| 25% | -2.72 | -4.20 | -2.68 | -5.64 |
| 50% | -0.14 | -0.66 | 1.50 | 0.27 |
| 75% | 2.61 | 3.35 | 5.90 | 6.41 |
| 95% | 7.72 | 10.71 | 17.72 | 22.83 |
| 99% | 12.37 | 17.86 | 35.32 | 42.52 |

**Table 4.4— Correlation coefficients for licensure tests**

| | CSET | CBEST | RICA |
|---|---|---|---|
| CSET | 1.00 | | |
| CBEST | 0.58 | 1.00 | |
| RICA | 0.44 | 0.46 | 1.00 |

**Table 4.5— Determinants of Teacher Unobserved Reading Heterogeneity in Levels Model**

| | ALL tests | CBEST | CSET | RICA |
|---|---|---|---|---|
| Years of teaching experience | 0.5366* | 0.5444* | 0.5620* | 0.5449* |
| | (0.1200) | (0.1009) | (0.1176) | (0.1092) |
| Teaching experience squared | -0.0347* | -0.0361* | -0.0360* | -0.0376* |
| | (0.0086) | (0.0070) | (0.0083) | (0.0082) |
| Female teacher | 0.4371 | 0.6641* | 0.5946* | 0.7345* |
| | (0.2759) | (0.2103) | (0.2526) | (0.2186) |
| Black/African American teacher | -0.1118 | 0.2646 | -0.1202 | 0.5172 |
| | (0.4609) | (0.4368) | (0.4501) | (0.4373) |
| Hispanic teacher | 0.1585 | 0.1187 | 0.1527 | 0.4169 |
| | (0.3136) | (0.2858) | (0.3022) | (0.2764) |
| Asian/Pacific Islander teacher | 0.5689 | 0.5248 | 0.5567 | 0.5946 |
| | (0.4130) | (0.3726) | (0.4087) | (0.3697) |
| Teacher has MA or Ph.D | -0.5091 | -0.9107 | -0.8205 | -0.9751 |
| | (0.5506) | (0.5520) | (0.5511) | (0.5722) |
| CBEST (standardized) | -0.4181* | -0.1531 | | |
| | (0.1747) | (0.1214) | | |
| CSET (standarized) | -0.0577 | | -0.1454 | |
| | (0.2059) | | (0.1469) | |
| RICA (standardized) | 0.2497 | | | 0.1056 |
| | (0.1767) | | | (0.1075) |
| Constant | -1.5235* | -1.6211* | -1.5956* | -1.7990* |
| | (0.3734) | (0.3352) | (0.3661) | (0.3422) |
| Adj.R-squared | 0.0179 | 0.0175 | 0.0172 | 0.0164 |
| Obs | 1981 | 2727 | 2033 | 2630 |

Note: Standard errors are in parenthesis. Standard errors are adjusted by the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

**Table 4.6— Determinants of Teacher Unobserved Math Heterogeneity in Levels Model**

| | ALL tests | CBEST | CSET | RICA |
|---|---|---|---|---|
| Years of teaching experience | 0.5433* | 0.5138* | 0.5893* | 0.5131* |
| | (0.1308) | (0.1158) | (0.1297) | (0.1186) |
| Teaching experience squared | -0.0330* | -0.0333* | -0.0360* | -0.0342* |
| | (0.0080) | (0.0071) | (0.0079) | (0.0074) |
| Female teacher | 0.0950 | 0.6945* | 0.3352 | 0.8227* |
| | (0.3754) | (0.3197) | (0.3545) | (0.3219) |
| Black/African American teacher | -0.3778 | -0.0762 | -0.4040 | 0.4464 |
| | (0.5139) | (0.4661) | (0.5054) | (0.4633) |
| Hispanic teacher | 0.3322 | 0.4185 | 0.3892 | 0.9285* |
| | (0.4134) | (0.3548) | (0.3948) | (0.3489) |
| Asian/Pacific Islander teacher | 0.8964 | 0.8677 | 0.8250 | 1.1622* |
| | (0.5760) | (0.4832) | (0.5684) | (0.4945) |
| Teacher has MA or Ph.D | -1.3310 | -1.8775* | -1.8666* | -2.1120* |
| | (0.7555) | (0.7780) | (0.7246) | (0.8124) |
| CBEST (standardized) | -0.6508* | -0.4353* | | |
| | (0.2193) | (0.1680) | | |
| CSET (standarized) | -0.3777 | | -0.4187* | |
| | (0.2193) | | (0.1743) | |
| RICA (standardized) | 0.5771* | | | 0.1961 |
| | (0.1850) | | | (0.1423) |
| Constant | -1.7059* | -1.9358* | -1.8276* | -2.2185* |
| | (0.4659) | (0.4136) | (0.4610) | (0.4166) |
| Adj.R-squared | 0.0217 | 0.0159 | 0.0183 | 0.0136 |
| Obs | 1981 | 2727 | 2033 | 2630 |

Note: Standard errors are in parenthesis. Standard errors are adjusted for the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

**Table 4.7— Determinants of Teacher Unobserved Reading Heterogeneity in Gains Model**

| | ALL tests | CBEST | CSET | RICA |
|---|---|---|---|---|
| Years of teaching experience | 1.2603* | 1.3478* | 1.3651* | 1.2964* |
| | (0.2109) | (0.2009) | (0.2193) | (0.1987) |
| Teaching experience squared | -0.0679* | -0.0697* | -0.0723* | -0.0742* |
| | (0.0111) | (0.0124) | (0.0115) | (0.0123) |
| Female teacher | 1.6442* | 1.3242* | 1.8706* | 1.5517* |
| | (0.5748) | (0.4915) | (0.5625) | (0.4901) |
| Black/African American teacher | -3.5697* | -3.2342* | -3.5759* | -2.6969* |
| | (1.1121) | (0.9788) | (1.1116) | (0.9545) |
| Hispanic teacher | 0.1141 | 0.1405 | 0.1502 | 0.7697 |
| | (0.6900) | (0.6367) | (0.6807) | (0.5845) |
| Asian/Pacific Islander teacher | 0.0792 | 0.3183 | 0.0091 | 0.3908 |
| | (0.9871) | (0.8834) | (0.9722) | (0.8588) |
| Teacher has MA or Ph.D | -2.4319* | -1.5259 | -1.8945 | -2.5685* |
| | (0.8856) | (1.0413) | (1.0973) | (0.8664) |
| CBEST (standardized) | -0.5197 | -0.4383 | | |
| | (0.4509) | (0.2750) | | |
| CSET (standarized) | -0.5345 | | -0.5563 | |
| | (0.4473) | | (0.3016) | |
| RICA (standardized) | 0.1974 | | | -0.0759 |
| | (0.3335) | | | (0.2579) |
| Constant | -1.0450 | -1.0012 | -1.3542 | -1.2711 |
| | (0.7895) | (0.6763) | (0.7785) | (0.6616) |
| Adj.R-squared | 0.0327 | 0.0288 | 0.0338 | 0.0277 |
| Obs | 1720 | 2334 | 1764 | 2261 |

Note: Standard errors are in parenthesis. Standard errors are adjusted for the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

**Table 4.8— Determinants of Teacher Unobserved Math Heterogeneity in Gains Model**

| | ALL tests | CBEST | CSET | RICA |
|---|---|---|---|---|
| Years of teaching experience | 1.3431* | 1.4617* | 1.4359* | 1.5388* |
| | (0.2447) | (0.2411) | (0.2426) | (0.2436) |
| Teaching experience squared | -0.0771* | -0.0806* | -0.0812* | -0.0943* |
| | (0.0132) | (0.0159) | (0.0133) | (0.0158) |
| Female teacher | 0.4468 | 1.6746* | 0.8004 | 1.6109* |
| | (0.7903) | (0.6982) | (0.7560) | (0.6908) |
| Black/African American teacher | -0.2708 | 0.4276 | -0.3353 | 0.8421 |
| | (1.0774) | (0.9952) | (1.0597) | (0.9215) |
| Hispanic teacher | 2.1242* | 2.6344* | 2.0805* | 3.2817* |
| | (0.8625) | (0.7933) | (0.8600) | (0.7730) |
| Asian/Pacific Islander teacher | 1.1129 | 1.2067 | 0.8768 | 1.5767 |
| | (1.1544) | (1.0509) | (1.1271) | (1.0689) |
| Teacher has MA or Ph.D | -1.5991 | -1.1347 | -1.5792 | -2.1641 |
| | (1.6483) | (1.5700) | (1.6104) | (1.5540) |
| CBEST (standardized) | -0.7001 | -0.4614 | | |
| | (0.5294) | (0.3561) | | |
| CSET (standarized) | -0.6967 | | -0.5959 | |
| | (0.4815) | | (0.3492) | |
| RICA (standardized) | 0.6925 | | | 0.2150 |
| | (0.3962) | | | (0.3152) |
| Constant | -2.5492* | -3.7835* | -2.7908* | -3.9725* |
| | (1.0369) | (0.9246) | (1.0014) | (0.9173) |
| Adj.R-squared | 0.0247 | 0.0253 | 0.0240 | 0.0268 |
| Obs | 1720 | 2334 | 1764 | 2261 |

Note: Standard errors are in parenthesis. Standard errors are adjusted for the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

**Table 4.9— Estimates of Value-Added models including only Student Fixed Effects**

| | Levels | | Gains | |
|---|---|---|---|---|
| Variable | Reading | Math | Reading | Math |
| Test year 2001 | 3.2368* | 2.9866* | | |
| | (0.0614) | (0.0689) | | |
| Test year 2002 | 5.7546* | 6.5570* | -2.5589* | -0.3029* |
| | (0.0902) | (0.1056) | (0.0928) | (0.1205) |
| Test year 2003 | 4.5058* | 6.0384* | -6.8583* | -5.5723* |
| | (0.1227) | (0.1451) | (0.1447) | (0.1900) |
| Test year 2004 | 5.5600* | 7.5748* | -2.3047* | -3.7603* |
| | (0.1453) | (0.1767) | (0.1993) | (0.2646) |
| Number of students in class | -0.0659* | -0.1081* | -0.0141 | 0.0230 |
| | (0.0060) | (0.0075) | (0.0103) | (0.0137) |
| Proportion female in class | -0.1574 | 0.5643* | -1.4785* | -0.6775 |
| | (0.2279) | (0.2786) | (0.4748) | (0.6109) |
| Proportion black/African American in class | -1.6568* | -2.4276* | -1.8692* | -2.3443* |
| | (0.3607) | (0.4182) | (0.7207) | (0.9222) |
| Proportion Hispanic in class | -1.6702* | -0.9215* | -1.9772* | -0.3996 |
| | (0.2840) | (0.3242) | (0.5764) | (0.7156) |
| Proportion Asian/Pacific Islander in class | -0.3901 | -0.3312 | -2.0639* | -0.5523 |
| | (0.4334) | (0.5070) | (0.8627) | (1.0885) |
| Hispanic student and teacher | 0.0336 | 0.5066* | 0.3888 | 0.7015* |
| | (0.1216) | (0.1486) | (0.2306) | (0.2892) |
| Black student and teacher | 0.4662* | 0.6290* | 0.4258 | 0.7051* |
| | (0.1428) | (0.1682) | (0.2594) | (0.2998) |
| Asian student and teacher | 0.0294 | 0.2122 | -0.7151* | -0.1041 |
| | (0.1690) | (0.1910) | (0.3524) | (0.4206) |
| Female student and teacher | -0.1599* | -0.2820* | -0.0904 | -0.0022 |
| | (0.0537) | (0.0576) | (0.1121) | (0.1224) |
| College parents and Teacher with MA/PhD | -0.1104 | -0.1341 | 0.2110 | 0.0121 |
| | (0.0929) | (0.1071) | (0.1683) | (0.2064) |
| Years of teaching experience | 0.0796* | 0.1092* | 0.1133* | 0.0873* |
| | (0.0091) | (0.0118) | (0.0164) | (0.0217) |
| Teaching experience squared | -0.0017* | -0.0024* | -0.0024* | -0.0015* |
| | (0.0003) | (0.0004) | (0.0005) | (0.0006) |
| Female teacher | 0.5405* | 0.3376* | 0.4317* | 0.1094 |
| | (0.0619) | (0.0821) | (0.1093) | (0.1463) |
| Black/African American teacher | -0.3291* | -1.0803* | -0.1056 | -1.0805* |
| | (0.0900) | (0.1147) | (0.1538) | (0.2026) |
| Hispanic teacher | 0.0850 | 0.0459 | 0.3705 | 0.3663 |
| | (0.1183) | (0.1454) | (0.2222) | (0.2816) |
| Asian/Pacific Islander teacher | 0.5750* | 0.9371* | 0.2611 | 0.9217* |
| | (0.0835) | (0.1146) | (0.1491) | (0.2031) |
| Teacher has MA or Ph.D | 0.1037 | 0.1197 | 0.1376 | -0.1425 |
| | (0.0693) | (0.0923) | (0.1243) | (0.1636) |
| CBEST (standardized) | -0.2250* | -0.3526* | -0.0946 | -0.1329 |
| | (0.0797) | (0.1097) | (0.1400) | (0.1888) |
| CSET (standardized) | -0.1296 | -0.2793 | -0.2509 | -0.4436 |
| | (0.1229) | (0.1906) | (0.2170) | (0.3403) |

| | | | | |
|---|---|---|---|---|
| RICA (standardized) | 0.1698* | 0.2601* | -0.0190 | 0.1049 |
| | (0.0702) | (0.0915) | (0.1111) | (0.1558) |
| Missing CBEST | -0.7017* | -0.5732 | -0.3372 | 0.8232 |
| | (0.2925) | (0.4595) | (0.6454) | (0.8077) |
| Missing CSET | 0.0550 | 0.2099 | 0.4478 | 0.3260 |
| | (0.1442) | (0.1960) | (0.2520) | (0.3560) |
| Missing RICA | 0.5016 | 0.2165 | 0.1769 | -1.3219 |
| | (0.2948) | (0.4611) | (0.6497) | (0.8174) |
| Constant | -1.3714* | -2.0306* | 4.9581* | 3.2992* |
| | (0.3171) | (0.3853) | (0.6219) | (0.8183) |
| Standard Deviation of Student Effects | 17.5693 | 19.4493 | 10.7753 | 12.0794 |
| R-squared | 0.0277 | 0.0302 | 0.0349 | 0.0195 |
| Obs | 935,775 | 935,775 | 585,325 | 585,325 |

Note: Standard errors are in parenthesis. Standard errors are adjusted for the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

**Table 4.10— Determinants of Student Unobserved Reading and Math Heterogeneity in Levels and Gains Models**

| | Levels | | Gains | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| Student is female | 2.4236* | -0.4602* | 0.5150* | 0.7927* |
| | (0.0677) | (0.0735) | (0.0443) | (0.0492) |
| Student is black | -13.3620* | -15.7509* | -1.1651* | -1.1970* |
| | (0.5005) | (0.5246) | (0.3427) | (0.3949) |
| Student is Hispanic | -6.5909* | -6.3081* | -0.0656 | -0.9982* |
| | (0.3205) | (0.3230) | (0.2912) | (0.3367) |
| Student is Asian/Pacific Islander | -1.0261* | 3.3837* | 0.2385 | -0.1005 |
| | (0.3860) | (0.4137) | (0.3130) | (0.3531) |
| LEP & Hispanic | -9.4199* | -7.2211* | 1.2652* | 0.1751 |
| | (0.1631) | (0.1864) | (0.1247) | (0.1463) |
| LEP & Asian/Pacific Islander | -7.2297* | -3.2146* | 2.6288* | 1.1075 |
| | (0.5443) | (0.6042) | (0.3187) | (0.6282) |
| LEP & other | -11.9889* | -6.3848* | 2.2163* | 0.3100 |
| | (0.5027) | (0.5341) | (0.4051) | (0.5787) |
| Student receives free/reduced lunch | -4.5619* | -4.1184* | 1.1669* | 0.3681 |
| | (0.2530) | (0.2488) | (0.1704) | (0.2181) |
| Parent is high school graduate | 2.2831* | 2.0539* | 0.0345 | -0.0519 |
| | (0.1167) | (0.1389) | (0.0937) | (0.1218) |
| Parent has some college | 4.1510* | 3.6633* | 0.2199 | 0.0291 |
| | (0.1530) | (0.1780) | (0.1232) | (0.1513) |
| Parent is college graduate | 6.4656* | 6.3065* | 0.5192* | 0.4065* |
| | (0.1919) | (0.2082) | (0.1423) | (0.1768) |
| Parent has some graduate training | 8.2979* | 7.8745* | 0.8740* | 0.5649 |
| | (0.3720) | (0.3996) | (0.2769) | (0.3077) |
| Parent education is missing | 1.1386* | 1.0684* | 0.5750* | 0.6263* |
| | (0.1959) | (0.2150) | (0.1518) | (0.2036) |
| Student is gifted | 19.9684* | 22.4162* | -1.5134* | -1.1764* |
| | (0.3252) | (0.3553) | (0.1637) | (0.1924) |
| Student in special education | -11.9275* | -14.0885* | 1.4635* | 0.9058* |
| | (0.2163) | (0.2439) | (0.1483) | (0.1704) |
| Constant | 11.5987* | 11.2889* | -2.2734* | -0.0753 |
| | (0.3795) | (0.4082) | (0.3534) | (0.3992) |
| Adj.R-squared | 0.2987 | 0.2482 | 0.0108 | 0.0027 |
| Obs | 332538 | 332538 | 325521 | 325521 |

Note: Standard errors are in parenthesis. Standard errors are adjusted by the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

## Appendix

**Table A-1: Determinants of Unobserved Teacher Heterogeneity Effect Controlling for Both Licensure Scores and Initial Test Failure**

| | Levels | | Gains | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| Years of teaching experience | 0.4840* | 0.5207* | 1.2504* | 1.3372* |
| | (0.1532) | (0.1759) | (0.2096) | (0.2416) |
| Teaching experience squared | -0.0363* | -0.0393* | -0.0679* | -0.0770* |
| | (0.0152) | (0.0163) | (0.0111) | (0.0131) |
| Female teacher | 0.3240 | -0.0074 | 1.5437* | 0.2760 |
| | (0.2733) | (0.3780) | (0.5965) | (0.8031) |
| Black/African American teacher | -0.1656 | -0.4451 | -3.7082* | -0.3813 |
| | (0.4652) | (0.5263) | (1.1204) | (1.0839) |
| Hispanic teacher | 0.1672 | 0.3069 | -0.0365 | 2.0587* |
| | (0.3178) | (0.4195) | (0.6857) | (0.8634) |
| Asian/Pacific Islander teacher | 0.5255 | 0.8489 | 0.0342 | 1.0901 |
| | (0.4165) | (0.5796) | (0.9919) | (1.1598) |
| Teacher has MA or Ph.D | -0.4486 | -1.2491 | -2.3930* | -1.5879 |
| | (0.5525) | (0.7587) | (0.8859) | (1.6437) |
| CBEST (standardized) | -0.4067* | -0.5563* | -0.3515 | -0.6238 |
| | (0.1705) | (0.2161) | (0.4686) | (0.5390) |
| CSET (standardized) | -0.1127 | -0.3618 | -0.5296 | -0.7858 |
| | (0.2010) | (0.2217) | (0.4502) | (0.4996) |
| RICA score (standardized) | 0.2089 | 0.6122* | 0.3398 | 0.7332 |
| | (0.1904) | (0.1945) | (0.3479) | (0.4105) |
| CBEST failure | 0.0189 | 0.2245 | 0.6908 | 0.8580 |
| | (0.2696) | (0.3527) | (0.6824) | (0.8145) |
| CSET failure | 0.1871 | 0.4672 | 0.4735 | 0.0889 |
| | (0.2747) | (0.3313) | (0.6345) | (0.8151) |
| RICA failure | -0.7481 | -0.2153 | 0.1419 | -1.1520 |
| | (0.4009) | (0.4798) | (0.8386) | (0.9702) |
| Constant | -1.2928* | -1.7556* | -1.3350 | -2.5592* |
| | (0.3921) | (0.4972) | (0.8487) | (1.0739) |
| Adj.R-squared | 0.0128 | 0.0189 | 0.0325 | 0.0247 |
| Obs | 1981 | 1981 | 1720 | 1720 |

Note: Standard errors are in parenthesis. Standard errors are adjusted for the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

**Table A-2: Determinants of Unobserved Teacher Heterogeneity Effect Controlling for Initial Test Failure and Not Licensure Score**

| | Levels | | Gains | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| Years of teaching experience | 0.4777* | 0.5308* | 1.2768* | 1.3722* |
| | (0.1501) | (0.1733) | (0.2087) | (0.2380) |
| Teaching experience squared | -0.0358* | -0.0408* | -0.0692* | -0.0789* |
| | (0.0148) | (0.0159) | (0.0110) | (0.0130) |
| Female teacher | 0.4865 | 0.3885 | 1.8220* | 0.7975 |
| | (0.2506) | (0.3557) | (0.5697) | (0.7909) |
| Black/African American teacher | 0.0523 | -0.0922 | -3.3265* | 0.2035 |
| | (0.4619) | (0.5103) | (1.1012) | (1.0740) |
| Hispanic teacher | 0.3457 | 0.5468 | 0.2004 | 2.4106* |
| | (0.3031) | (0.4092) | (0.6567) | (0.8588) |
| Asian/Pacific Islander teacher | 0.5783 | 0.8480 | 0.0991 | 1.1409 |
| | (0.4161) | (0.5796) | (0.9873) | (1.1561) |
| Teacher has MA or Ph.D | -0.6263 | -1.5555* | -2.6226* | -1.9953 |
| | (0.5605) | (0.7562) | (0.9164) | (1.6282) |
| CBEST failure | 0.1135 | 0.2989 | 0.8704 | 1.0551 |
| | (0.2622) | (0.3452) | (0.6575) | (0.7956) |
| CSET failure | 0.3503 | 0.6292 | 0.7039 | 0.3861 |
| | (0.2618) | (0.3341) | (0.5759) | (0.7519) |
| RICA failure | -0.6780 | -0.1841 | 0.2787 | -1.0121 |
| | (0.3875) | (0.4589) | (0.8082) | (0.9211) |
| Constant | -1.5348* | -2.1625* | -1.8003* | -3.2523* |
| | (0.3689) | (0.4732) | (0.7704) | (1.0042) |
| Adj.R-squared | 0.0103 | 0.0118 | 0.0322 | 0.0225 |
| Obs | 1981 | 1981 | 1720 | 1720 |

Note: Standard errors are in parenthesis. Standard errors are adjusted for the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level.

**Table A-3: Determinants of Unobserved Teacher Heterogeneity with Nonlinear Licensure Effects**

| | Levels | | Gains | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| Years of teaching experience | 0.5234* | 0.5412* | 1.2551* | 1.3747* |
| | (0.1207) | (0.1293) | (0.2089) | (0.2427) |
| Teaching experience squared | -0.0342* | -0.0333* | -0.0675* | -0.0791* |
| | (0.0087) | (0.0079) | (0.0109) | (0.0134) |
| Female teacher | 0.4981 | 0.2408 | 1.771* | 0.7384 |
| | (0.2598) | (0.3734) | (0.5655) | (0.7653) |
| Black/African American teacher | -0.0461 | -0.2179 | -3.4653* | -0.0644 |
| | (0.469) | (0.5047) | (1.1318) | (1.0928) |
| Hispanic teacher | 0.2253 | 0.4417 | 0.1672 | 2.2095* |
| | (0.3078) | (0.4111) | (0.6873) | (0.8805) |
| Asian/Pacific Islander teacher | 0.5773 | 0.926 | 0.1069 | 1.1094 |
| | (0.4112) | (0.5745) | (0.977) | (1.1684) |
| Teacher has MA or Ph.D | -0.6148 | -1.544* | -2.6213* | -1.8125 |
| | (0.5532) | (0.753) | (0.8861) | (1.6502) |
| Medium CBEST | -0.2475 | -0.5572 | -0.2207 | -0.0282 |
| | (0.2972) | (0.3772) | (0.7735) | (0.8629) |
| High CBEST | -0.5739 | -1.3081* | -0.7134 | -1.8549 |
| | (0.3707) | (0.4516) | (1.0399) | (1.0717) |
| Medium CSET | -0.1873 | -0.5652 | -0.7335 | -0.9665 |
| | (0.2744) | (0.3541) | (0.6786) | (0.8125) |
| High CSET | -0.2898 | -0.5375 | -1.324 | -0.2133 |
| | (0.3576) | (0.4474) | (1.0388) | (0.9927) |
| Medium RICA | 0.4636 | 0.462 | -0.8441 | -0.2686 |
| | (0.2745) | (0.3519) | (0.6567) | (0.8015) |
| High RICA | 0.3575 | 0.9751* | 0.1327 | 0.7586 |
| | (0.3337) | (0.4156) | (0.7192) | (0.9528) |
| Constant | -1.3409* | -1.1739* | 0.2507 | -1.7761 |
| | (0.4577) | (0.5889) | (1.0021) | (1.2665) |
| Adj.R-squared | 0.0219 | 0.0238 | 0.0388 | 0.0302 |
| Obs | 1981 | 1981 | 1720 | 1720 |

Note: Standard errors are in parenthesis. Standard errors are adjusted the fact that teachers are clustered within schools. An asterisk indicates significance at a 95% level. Licensure variables are relative to lowest licensure category which is a score in the lowest 40$^{th}$ percentile of the test. The medium group indicates a licensure score from the 40$^{th}$ percentile to the 75$^{th}$ percentile. The high group indicates a licensure score at the 75$^{th}$ percentile or above.