



Munich Personal RePEc Archive

How effective are Los Angeles elementary teachers and schools?

Buddin, Richard

RAND Corporation

31 August 2010

Online at <https://mpra.ub.uni-muenchen.de/27366/>
MPRA Paper No. 27366, posted 10 Dec 2010 22:36 UTC

How Effective Are Los Angeles Elementary Teachers and Schools?

Richard Buddin

August 2010

1. INTRODUCTION

Over the past three decades, educational reforms have played a major role in local, state, and federal policy debates. Several factors drive the push for reforms: U.S. students have not performed as well on science and math tests as students in other industrial countries. Low-income and minority students lag significantly behind other students on most measures of academic achievement. The private sector has pushed for improvements in human capital investment, so the U.S. can remain competitive in the growing high-technology economy.

In 1983, the landmark publication *A Nation at Risk* voiced broad concerns that the U.S. was not providing education adequate for the needs of the 21st century.¹ States developed curriculum standards and standardized achievement tests to measure the performance of schools and districts. The No Child Left Behind (NCLB) Act of 2001 added federal goals and performance targets for schools and districts. NCLB also required a “highly qualified teacher” in all classrooms and public reporting of teacher qualifications. In 2009, the federal “Race to the Top” initiative encouraged states to develop rigorous student achievement standards and to use student achievement in teacher assessments. In addition to these federal initiatives, state and local governments have improved accountability and taken steps to improve student achievement.

In the past several years, new research has emerged that more effectively measures student achievement from year to year and creates the potential to tie student progress with individual teachers and other school inputs. NCLB and a series of other reforms have led states to develop annual testing of students in most grades (at least in reading and mathematics). Several states and some districts maintain individual student identifiers that allow researchers to track student progress from year to year and link that progress with changes in school resources or teachers or school practices. This type of data collection offers researchers improved tools for measuring how individual school inputs affect student outcomes.

The new measures rely on so-called “value-added” methods that isolate the contribution of a teacher or school to student learning conditional on individual student background and preparation. Teachers or schools are characterized as “high quality” if their students

Note: Richard Buddin is a senior economist at the RAND Corporation. He performed this study as an independent contractor for the *Los Angeles Times*. The RAND Corporation was not involved in the study or analysis. He is grateful to Gema Zamarro of RAND and several anonymous reviews for their comments on the study.

¹ National Commission of Excellence in Education (1983)

make above average improvements in student achievement relative to other teachers or schools with comparable students.

The new methods are contrasted with traditional metrics that focus on the average achievement level of students at a school. In the traditional approach, "high-performing" schools have students with higher achievement or proficiency levels than the average school. The problem with this approach is that student achievement is strongly influenced by student background and preparation. As a result, the "high-performing" schools are inevitably schools with few disadvantaged or low-income students. Traditional methods are ill suited for separating a school's success in improving student outcomes from their success in attracting students with strong preparation. By design, value-added measures isolate whether some schools (or teachers) do a better job of *improving* student achievement than do others.

This study focuses on value-added measures of elementary school student achievement in the Los Angeles Unified School District (LAUSD). LAUSD is the second largest school district in the United States (behind New York City Public Schools) with about 700,000 students and 35,000 teachers. LAUSD, like many large urban districts, has large shares of low-income and minority students. About 84 percent of students are eligible for free/reduced school lunch, and almost 40 percent come from families where neither parent completed high school. About 76 percent of students are Hispanic and another 9 percent are black. Nearly half of elementary students are English language learners (ELLs) and receive special instruction to improve their English proficiency. Many at-risk students are concentrated in some schools and neighborhoods, so this isolation means these students have little interaction with more affluent peers.

This study addresses three key issues:

1. How much does teacher quality vary from school to school and from teacher to teacher?
2. What teacher qualifications or backgrounds are associated with teacher success in the classroom?
3. How do traditional measures of school performance compare with value-added measures of teacher and school effectiveness?

We will rely on student-level longitudinal data to track individual student progress from year to year and to identify the teachers and schools that are most effective at improving student achievement.

At the outset, we acknowledge several limitations of value-added measures. First, student achievement tests are not administered until 2nd grade, so the measures provide no indication of the effectiveness of kindergarten or 1st grade teachers.² Second, annual tests are only given in English Language Arts (ELA) and math. These subjects are important

² An empirical concern is that if a school is particularly effective in teaching kindergarten and 1st grade students, then they may have less potential to improve the outcomes for students in 2nd through 5th grades. Alternatively, schools with poor kindergarten and 1st grade preparations may set the stage for strong performance in 2nd through 5th grades. In our analysis, we implicitly assume that school elementary school performance is relatively consistent across grades.

and key building blocks for other subjects, but the tests do not provide a comprehensive indication of what students learned or everything that they should know at their grade level. Third, standardized tests are imperfect measures of learning because students may misunderstand what is expected or because individual students may have test anxiety or other issues on the day of the test. Some of these problems will "average" out across students in a classroom or school.

While these deficiencies in value-added models are a concern, the models offer an important opportunity to identify what factors improve student outcomes overtime. Value-added approaches are not intended to replace measures of student proficiency as indications of academic success, but the new approaches offer valuable insights into how districts might align their resources to improve the proficiency levels of all students.

The remainder of this paper is divided into four sections. Section 2 provides a description of the value-added framework and the statistical models used in this study. Section 3 describes the characteristics of the students and teachers in LAUSD and shows how these factors vary with traditional measures of school performance. Section 4 reports value-added estimates of teacher and school effectiveness. The section compares these estimates with traditional measures of school performance. The final section offers conclusions and recommendations.

2. STATISTICAL APPROACH

An education production function is the underlying basis for nearly all recent studies of student achievement (Buddin and Zamarro, 2009). 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. Following Todd and Wolpin (2003), let T_{it} be the test score measure of student i that is observed in year t and ε_{it} be a measurement error, and let X_{it} and v_{it} represent observed and unobserved inputs for student i at time t . Finally, let μ_{i0} be the student's endowed ability that does not vary over time. Assume that the cognitive production function is linear in the inputs and in the unobserved endowment and that input effects do not depend on the child's age but may depend on the age at which they were applied relative to the current age. Then, a general cognitive production function will be given by:

$$T_{it} = \mu_{i0} + \alpha_1 X_{it} + \alpha_2 X_{it-1} + \dots + \rho_1 v_{1t} + \rho_2 v_{1t-1} + \dots + \varepsilon_{it} \quad (1)$$

where test scores in a given year are a function of current and past observed and unobserved inputs as well as of the initial ability of the child.

Estimation of Equation 1 requires a comprehensive history of all past and present family and school/teacher inputs as well as information about each student's endowed ability. Several empirical problems complicate the estimation of this complete, ideal model:

- Endowed ability (μ_{i0}) or some student inputs are not observed, and observed student inputs may be chosen endogenously with respect to them (student unobserved heterogeneity). For example, English learner status (an observed variable) may be correlated with family wealth (an unobserved variable). If so, the

estimated effect of English learner status may reflect the underlying wealth effect in addition to the direct effect of being an English learner.

- Data sets on teacher inputs are incomplete, and observed teacher inputs may be chosen endogenously with respect to the unobserved teacher inputs (teacher unobserved heterogeneity). For example, teacher effort may be difficult to measure, and effort might be related to measured teacher qualifications, i.e., teachers with higher licensure test scores may regress to the mean with lower effort.
- Students and teachers are not allocated randomly into schools or classrooms. Families with higher preferences for schooling will try to allocate their children in better schools or classrooms, principals may not allocate teachers to classrooms randomly, and good teachers may have more negotiation power to locate themselves in schools or classrooms with higher-achieving students. These choices will lead to endogeneity of observed inputs with respect to unobserved student and teacher inputs or endowments.

Different specifications have been proposed in the most recent literature to try to overcome previous data limitations (Buddin and Zamarro, 2009).

Measuring Teacher Quality

In this paper, we start with a general dynamic panel data model that includes student and teacher fixed effects in the following reduced form:

$$T_{it} = T_{it-1} \lambda + x_{it} \beta_1 + u_i \eta + q_j \rho + \alpha_i + \phi_j + \varepsilon_{it} \quad (2)$$

where T_{it} is either the English Language Arts (ELA) or math test score for 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 aptitudes and needs); and q_j are time-invariant observable characteristics of the j^{th} teacher (gender, education, experience), and λ indicates the persistence of prior-year learning. The model includes individual student and teacher fixed effects (α_i and ϕ_j). Finally, ε_{it} contains individual and teacher time variant unobserved characteristics.^{3,4}

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 identified only by the students who change teachers. It is assumed that ε_{it} is strictly exogenous. That is, student's assignments to teachers are independent of ε_{it} . Note, according to this assumption, assignment of students to teachers may be a function of the observables and the time-invariant unobservables.

³ We discuss modeling issues in more detail in our earlier paper on student achievement in elementary school (See Buddin and Zamarro, 2009).

⁴ We also estimated fixed effects levels model assuming $\lambda=0$ and a gains model assuming $\lambda=1$. We prefer the more general model in Eq. 2, because it incorporates a more flexible adjustment for student heterogeneity. The teacher effects from the dynamic panel model are similar to those for the more restrictive levels and gains models (Buddin and Zamarro, 2009)..

The model was simplified by assuming that the student heterogeneity term (α_i) was zero. This assumption was consistent with initial data runs that indicated that student heterogeneity was statistically insignificant after controlling for prior year test score and observed student characteristics. More importantly, recent research has shown that this type of model performs well in predicting teacher performance from year to year in both experimental and non-experimental settings (Kane and Staiger, 2008; McCaffrey et al., 2009).

Teacher heterogeneity (ϕ_j) is probably correlated with observable student and teacher characteristics (e.g., non-random assignment of students to teachers). Therefore, random effect methods are inconsistent, and the fixed teacher effects are estimated in the model. The fixed teacher effects are defined as $\psi_j = \phi_j + q_j \rho$.

The model is estimated in two steps. In a first step, we estimate the following equation using fixed teacher effects:

$$T_{it} = T_{it-1} \lambda + x_{it} \beta_1 + u_i \eta + \psi_j + \varepsilon_{it} \quad (3)$$

In the second stage, we evaluate how individual teacher characteristics affect value-added estimates of teacher quality (ψ_j). Many of the observable teacher characteristics considered in this analysis are important determinants of teacher recruitment, retention and salary decisions.

Causal interpretation of the coefficients in these second step regressions would need the additional assumptions that $\text{Cov}(q_j, \phi_j) = 0$. As explained below, this assumption is unlikely to be satisfied in this context. Thus, our second step estimates should not be interpreted as causal effects but as measures of the correlation between observed characteristics and the teacher quality and student ability terms.

We used an instrumental variable approach for dealing with possible measurement error in lagged student achievement. The lagged math score was used as an instrument for the lagged ELA score in the ELA achievement equation. Similarly, the lagged ELA score was used as an instrument for the lagged math score in the math achievement equation. This approach reduces some noise in the prior year test score and improves the quality of the model estimates.

Finally, our dependent variable in these second step regressions is a statistical estimate of the true measures of teacher quality (ψ_j), so it is 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).

Measuring School Quality

The school quality model is a slight variant of the teacher quality model, where teacher fixed effects are replaced with school fixed effects:

$$Y_{it} = Y_{it-1} \lambda + x_{it} \beta_1 + u_i \eta + \tau_k + \varepsilon_{it} \quad (4)$$

where τ_k is the value-added measure of school quality at the k^{th} school attended by the student.

In principal, the models could be combined with both teacher and school effects. We observe relatively few teachers switching schools over time, however, so it is difficult to identify separate teacher and school effects in a combined model.

The school fixed effects could be decomposed into various elements in a second stage regression as proposed for teacher fixed effects. We have relatively few school characteristics, except the mix of students and teachers at each school. As a result, our school quality model focuses on the estimation of Eq. 4 with no further second stage analysis of school factors that contribute to school quality. Student-level factors are implicitly included in Eq. 4 through the current and lagged test scores. Teacher-level factors are modeled in the teacher quality model.

3. DATA

Our district data covers the 2002-2003 through 2008-2009 school years. The elementary school analysis is based on about 1.5 million student/year records for students enrolled in grades 2 through 5 over the seven year period. Elementary students are not tested in kindergarten or first grade. We observe student test scores for about 603,500 different elementary students taught by about 18,000 teachers in 520 schools.

Table 1 shows the characteristics of students in LAUSD elementary schools and describes how these characteristics vary with overall achievement at those schools. Achievement is measured by the California Academic Performance Index (API), a school-level measure of student test performance on the California Standards Test (CST). The CST is aligned with state curriculum standards and administered to nearly all students in grades 2 through 11 each spring.

LAUSD has large percentages of Hispanic students, English Learners, and students from low income families. About 76 percent of students are Hispanic. English Language Learners (ELLs) comprise 39 percent of enrollments with another 15 percent of students Reclassified Fluent-English Proficient (RFEP). RFEP means that those students were not initially English proficient when they entered school but subsequently became proficient while in school. About 84 percent of students receive free/reduced school lunch, and 37 percent of students come from families where neither parent graduated from high school. About 10 percent of students are in gifted programs, while another 10 percent have some type of disability.

Student characteristics vary widely between low-API and high-API schools. The high-API schools are comprised of a smaller share of Hispanics, a larger share of Asians, and smaller shares of ELLs and RFEPs than are the low-API schools. Family wealth differs substantially across these schools as well. While 55 percent of students in the top quartile school are eligible for free/reduced lunch, nearly all students (97 percent) in the lowest quartile are eligible. About 50 percent of parents from low-API schools have not

completed high school as compared with only 16 percent of parents for high-API schools. The percentage of students with disabilities is invariant across schools in different API quartiles.

Table 1: Characteristics of Elementary School Students by School API in 2009

Student Characteristic	1st API	2nd API	3rd API	4th API	Overall
	(Lowest) Quartile			(Highest) Quartile	
Black	12	8	9	8	9
Asian	1	1	3	10	4
Hispanic	87	87	80	50	76
Free/Reduced Lunch	97	94	89	55	84
Gifted	4	5	7	19	9
English Learner	51	47	39	20	39
Reclassified Fluent-English-Proficient	16	17	17	10	15
Disabilities	11	11	11	11	11
<i>Parents Education Level</i>					
Not High School Grad	50	44	37	16	37
High School Grad	30	30	30	21	28
Some College	13	16	19	24	18
College Graduate	5	8	10	26	12
Some Graduate School	2	3	4	14	6

Teacher characteristics differ much less across elementary schools than do student characteristics. Table 2 shows that teacher experience and education level differ little between low- and high-API schools. Nearly all teachers have full teaching credentials. Black and Hispanic teachers are much more likely to work in low- than high-API schools. Finally, most elementary teachers are women, but the share rises from 72 percent for low-API schools to 82 percent for high-API schools.

Table 2: Characteristics of Elementary School Teacher by School API in 2009

Teacher Characteristics	1st API	2nd API	3rd API	4th API	Overall
	(Lowest) Quartile			(Highest) Quartile	
Black	13	9	9	6	10
Hispanic	50	48	40	23	41
Asian/Pacific Islander	6	7	9	16	9
Female	72	75	76	82	76
Masters or PhD	32	30	32	30	31
Years of Experience	12.4	13.1	12.8	13.2	12.8
Experience < 4 Years	10	8	8	9	9
Full Credential	100	99	99	99	99

Student test scores vary considerably from school to school, as expected from the large differences in student backgrounds across LAUSD. Table 3 shows that only about 25

percent of students are ELA proficient or above in low-API schools as compared with 60 percent for the high-API group. The proficiency levels are much higher in math than in ELA, but the gap between the low- and high-API groups is about the same. Even the high-API schools are well short of the state and national goals of having all students proficient at their grade level.

Table 3. Test Score Performance by School API in 2009

	1st API (Lowest) Quartile	2nd API Quartile	3rd API Quartile	4th API (Highest) Quartile	Overall
API	682	730	768	852	759
ELA proficient or above	25	33	39	60	40
Math proficient or above	41	50	57	72	55

Table 3 also provides an underlying indication of the distribution of API scores across schools. The interquartile gap between the 1st and 2nd quartiles is 48 points as compared with a gap of 84 between the 3rd and 4th quartile. This difference reflects that fact that many schools are clustered around relatively low API scores, while a few schools score considerably higher than the mean.

4. RESULTS

Teacher Effectiveness

The results of the 1st stage regression estimates for ELA and math achievement (Equation 3) are shown in Table 4. The student’s raw scores in ELA and math are standardized by grade and year.

The results show strong persistence of achievement from one year to the next, i.e., λ is about 0.87 in both equations. The proximity of λ to unity suggests that the results of the lagged achievement model will be similar to that of the gains model where λ is restricted to equal one.

Several student-level characteristics have a significant effect on achievement even after conditioning on the student’s test score in the prior year. The effect sizes of the student variables are small, however. Title I participants have scores lower than other students with an effect size of 0.03 in ELA and 0.05 in math. Girls have achievement scores about 0.02 standard deviations higher than do comparable boys. English language learners (ELLs) do worse than other students in both ELA and math, but the effect size in ELA is more than two times as large as in math. Finally, the data show whether students started school in LAUSD or joined the district after kindergarten. The results show that late joiners do better than students starting in the district.

The grade and year variables in Table 4 are control factors. They adjust for differences in test results from year to year and from grade to grade.

**Table 4: ELA & Math Achievement Regressions
for Teacher Effectiveness**

	ELA	Math
Lagged ELA	0.8762* (0.0011)	
Lagged Math		0.8709* (0.0011)
Grade 4	0.0130* (0.0029)	0.0059 (0.0032)
Grade 5	0.0260* (0.0031)	0.0171* (0.0034)
Title I	-0.0316* (0.0033)	-0.0533* (0.0037)
Female	0.0267* (0.0012)	0.0212* (0.0013)
English Language Learner	-0.0272* (0.0015)	-0.0117* (0.0016)
Joined after Kindergarten	0.0262* (0.0013)	0.0204* (0.0014)
Test Year 2005	0.0140* (0.0020)	0.0156* (0.0022)
Test Year 2006	0.0051* (0.0021)	0.0053* (0.0024)
Test Year 2007	0.0058* (0.0022)	0.0039 (0.0024)
Test Year 2008	0.0070* (0.0023)	0.0069* (0.0025)
Test Year 2009	-0.0028 (0.0024)	0.0019 (0.0026)
Constant	0.0175* (0.0039)	0.0342* (0.0042)
Teacher Effects (σ_{ψ})	0.2101	0.2968
R-squared	0.6853	0.5960
Student Years	836310	836310
Teachers	11503	11503

* Statistically significant at 5% confidence Level. Robust standard errors are in parentheses. The omitted reference categories are grade 3, not in a Title I school, male, not an ELL, joined LAUSD in kindergarten, and test year 2004. The dependent variables are student ELA and math test scores standardized by grade and year. The F test that all $\psi_j=0$ is 9.06 in ELA and 16.76 in math.

The table shows the standard deviations of the teacher effects in ELA and math. We used Bayesian methods to shrink these estimates and correct for measurement error. The adjusted effect sizes are 0.1813 ELA and 0.2678 in math. These effects sizes are large and suggest that students assigned to “high” quality teachers have much higher test scores at the end of the year than students assigned to “low” quality teachers. For example, a typical student moves from the 50th ELA percentile with an average teacher to the 57th

percentile for a teacher one standard deviation above the average. The gap in math is larger, where the student moves from the 50th math percentile with the average teacher to the 61st percentile for a teacher one standard deviation above the average.

How large are differences in teacher quality?

Hill et al. (2008) provide criteria for judging the importance of various factors affecting student achievement. They discuss three types of benchmarks: 1) average annual growth in achievement, 2) gaps in student achievement by demographic groups, and 3) gains of educational interventions.

In nationally normed tests, Hill et al. (2008) measured the effect size of learning gains from year to year. They found gains were larger in early grades as compared with higher grades and were higher in math than in ELA. At the 4th grade level, the average annual achievement gain has an effect size of 0.40 in ELA and 0.56 in math. The estimates from LAUSD show teacher effect sizes that are about half of the national average learning gains in ELA and math.

The second benchmark looks at the academic achievement gaps for disadvantaged groups. At the 4th grade level, Hill et al. (2008) report an ELA effect size gap between black/white students of -0.83, between Hispanic/white students of -0.77, and between eligible/ineligible students in free/reduced lunch program of -0.74. The gap is about 0.10 higher in math effect sizes.

The achievement gaps are three to four times larger than the estimated teacher effects for LAUSD. In an earlier study, Gordon, Kane, and Staiger (2006) found similarly large differences in teacher effectiveness. They argued that minority students assigned to a top quartile teacher four years in a row instead of a bottom quartile teacher would close the test-score gap.

The third benchmark looks at the effect sizes of school reforms implemented at the elementary school level. The average effect size of these interventions is 0.33 (Hill et al., 2008). By this third metric, as with the two earlier ones, the magnitude of estimated teacher effects is large.

How do teacher qualifications and background affect teacher value added?

Teacher experience and educational background have weak effects on teacher effectiveness (see Tables 5).⁵ Teacher experience has little effect on ELA scores beyond the first couple years of teaching--teachers with less than 3 years of experience gave teacher effects 0.05 standard deviations lower than comparable other teachers with 10 or more years of experience. Students with new teachers score 0.02 standard deviations lower in math than with teachers with 10 or more years of experience, but the effect is not statistically different from zero. These effect sizes mean that students with the most experienced teachers would average 1 or 2 percentile points higher than a student with a

⁵ The regressions in Table 5 are based on teachers that had at least 30 students with valid test scores.

new teacher. These effects are small relative to the benchmarks established by Hill et al. (2008).

Table 5. ELA and Math Teacher Effects and Teacher Characteristics

	ELA	Math
Experience < 3 years	-0.0504* (0.0072)	-0.0197 (0.0106)
Experience 3-5 years	-0.0117* (0.0058)	0.0212* (0.0091)
Experience 6-9 years	-0.0019 (0.0053)	0.0207* (0.0076)
Bachelor's + 30 semester hours	-0.0037 (0.0051)	-0.0030 (0.0076)
Master's	0.0129 (0.0076)	0.0016 (0.0108)
Master's + 30 semester hours	0.0061 (0.0072)	0.0090 (0.0104)
Doctorate	-0.0263 (0.0221)	-0.0380 (0.0342)
Full Teaching Credential	0.0048 (0.0092)	0.0092 (0.0156)
Black/African American	-0.0478* (0.0086)	-0.0745* (0.0114)
Hispanic	-0.0074 (0.0048)	-0.0058 (0.0072)
Asian/Pacific Islander	0.0343* (0.0071)	0.0739* (0.0105)
Female	0.0432* (0.0046)	0.0211* (0.0070)
Grade 4	-0.0091 (0.0066)	-0.0057 (0.0093)
Grade 5	-0.0144* (0.0059)	-0.0142 (0.0090)
Constant	-0.0185 (0.0114)	-0.0257 (0.0194)
R-squared	0.0274	0.0195
N	8719	8719

Notes: The dependent variables are ELA and math teacher effects from stage one, adjusted for measurement error. * indicates $p < 0.05$. Robust standard errors are in parentheses. The omitted categories are White non-Hispanics, Male, BA only, no full teaching credential, experience of 10 or more years, and grade 3.

Other teacher qualifications have little effect on student achievement. Teacher education beyond a bachelor's degree has no statistically significant effect on ELA or math achievement. Similarly, teachers with full teaching credentials are no more successful at improving student achievement than are teachers without credentials.

Teacher demographics have some effect on student achievement. Black/African American teachers have student gains about 0.05 and 0.07 standard deviations lower in ELA and math, respectively, than those of white non-Hispanic teachers. Asian/Pacific Islander teachers do better than their white non-Hispanic counterparts with effect sizes of 0.03 in ELA and 0.08 in math. Hispanic teachers have comparable outcomes with white non-Hispanic teachers. Female teachers have higher gains than comparable male teachers with an effect size of 0.04 in ELA and 0.02 in math.

Some teachers have suggested that students are more successful in some grades than others. The statistical approach examines student progress by year and by grade, however, so the estimated teacher effects should provide little advantage to teachers in some grades. The results in Table 5 show that grade effects are generally insignificant with only a small (0.01) effect for grade 5 ELA teachers relative to grade 3 ELA teachers.

The small and generally insignificant effects of teacher qualifications are consistent with several recent studies of teacher value added. Using Texas data, Rivkin et al. (2005) found that teacher experience and education explained a small share of the differences in teacher effectiveness across classrooms. Jacob and Lefgren (2008) also found weak effects of teacher qualifications on teacher quality. Harris and Sass (2006) found small effects of teacher experience and background on teaching effectiveness in Florida. Aaronson et al. (2008) found strong effects of Chicago teachers on achievement, but traditional measures of teacher quality like experience, education, and credential types had little effect on classroom results. Finally, Koedel and Betts (2007) looked at elementary school students in San Diego and found that traditional teacher quality measures (experience, quality of undergraduate college, education level, and college major) had little effect on student achievement.

In contrast with these studies, Clotfelter et al. (2007) did find positive effects of teacher experience, education, and teaching credentials for achievement in North Carolina elementary schools. The authors show that bundling of teacher qualifications does produce effect sizes differences of about 0.20 in math and 0.12 in reading. As a result, these authors argue that traditional measures of teacher quality do have an important effect on student achievement.

The key difference between the North Carolina results and our LAUSD results is that we have much smaller effects of traditional teacher qualifications on student achievement and those effects are often insignificantly different from zero. Table 5 shows that experience is the only qualification that is statistically significant and most of the experience effect is for new ELA teachers.

Table 6 shows how a teacher's grade, class size, and the prior achievement of their students affect teacher effectiveness. In California, class size differs sharply between 3rd and 4th grade due to state mandated class size limits through grade 3. Between 2002 and 2009, the average class size for 3rd grade was 19 as compared with 28 for 5th and 6th grades. The table show separate estimates of teacher effectiveness by grades to isolate possible class size effects over these grades.

Table 6. ELA and Math Teacher Effects, Teacher Characteristics, and Classroom Composition by Grade

	ELA			Math		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
Experience < 3 years	-0.0604* (0.0158)	-0.0297* (0.0119)	-0.0199 (0.0102)	-0.0327 (0.0228)	0.0049 (0.0175)	0.0168 (0.0182)
Experience 3-5 years	-0.0260* (0.0125)	0.0100 (0.0091)	0.0053 (0.0076)	-0.0174 (0.0162)	0.0478* (0.0138)	0.0577* (0.0156)
Experience 6-9 years	-0.0003 (0.0103)	0.0054 (0.0088)	0.0084 (0.0071)	0.0114 (0.0128)	0.0281* (0.0127)	0.0468* (0.0135)
Bachelor's + 30 sem. hrs.	-0.0052 (0.0096)	-0.0049 (0.0080)	0.0024 (0.0075)	-0.0106 (0.0122)	0.0193 (0.0126)	-0.0135 (0.0133)
Master's	0.0149 (0.0145)	-0.0112 (0.0124)	0.0220* (0.0094)	0.0028 (0.0186)	-0.0105 (0.0185)	0.0008 (0.0188)
Master's + 30 sem. hrs.	0.0195 (0.0145)	0.0022 (0.0105)	0.0030 (0.0096)	0.0122 (0.0177)	0.0047 (0.0157)	0.0194 (0.0175)
Doctorate	0.0093 (0.0544)	-0.0531 (0.0309)	-0.0344 (0.0260)	0.0140 (0.0656)	-0.0430 (0.0523)	-0.0791 (0.0567)
Full Teaching Credential	-0.0204 (0.0245)	0.0095 (0.0150)	0.0058 (0.0139)	0.0206 (0.0326)	-0.0216 (0.0222)	0.0227 (0.0269)
Black/Afr. Amer.	-0.0437* (0.0162)	-0.0426* (0.0131)	-0.0111 (0.0091)	-0.0963* (0.0187)	-0.0616* (0.0194)	-0.0203 (0.0178)
Hispanic	0.0132 (0.0092)	0.0274* (0.0079)	0.0104 (0.0069)	-0.0026 (0.0126)	0.0234 (0.0120)	0.0094 (0.0123)
Asian/Pac. Isl.	0.0428* (0.0122)	0.0252* (0.0119)	0.0231* (0.0099)	0.0864* (0.0165)	0.0549* (0.0176)	0.0587* (0.0185)
Female	0.0591* (0.0097)	0.0321* (0.0074)	0.0141* (0.0063)	0.0403* (0.0131)	0.0207 (0.0119)	-0.0157 (0.0115)
Average Lagged Test Score	0.0286* (0.0051)	0.0459* (0.0042)	0.0297* (0.0034)	0.0112 (0.0064)	0.0350* (0.0060)	0.0328* (0.0059)
Class Size	-0.0015 (0.0017)	0.0017 (0.0013)	0.0026* (0.0011)	-0.0008 (0.0025)	0.0002 (0.0020)	0.0033 (0.0018)
Constant	0.0146 (0.0428)	-0.0896* (0.0408)	-0.1084* (0.0344)	-0.0258 (0.0606)	-0.0358 (0.0596)	-0.1484* (0.0578)
R-squared	0.0426	0.0855	0.0570	0.0267	0.0365	0.0300
N	3226	2827	2666	3226	2827	2666

Notes: The dependent variables are math teacher effects from stage 1, adjusted for measurement error. * indicates $p < 0.05$. Robust standard errors are in parentheses. The omitted categories are White non-Hispanics, Male, BA only, no full teaching credential, experience of 10 or more years, and grade 3.

The patterns for individual grades show that class size is unrelated to teacher effectiveness in ELA and math for each grade. The results do show that teachers with better prepared students have some small advantage in measured effectiveness. A one standard deviation in the mean ELA and math scores of a teacher's new students is associated with about a 0.03 increase in the teacher's value added. Teachers gain some advantage if they are assigned better students, but the edge in measured teacher effectiveness is small.

Table 6 also shows the effects of traditional measures of teacher qualifications by grade. The results are consistent with the cross-grade results in Table 5. Inexperienced ELA teachers are less effective than more experienced teachers, but teachers with even three years of experience are nearly as effective as more veteran teachers. In math, the evidence shows that 4th grade teachers with 3 to 9 years of teaching are more effective than either new or more veteran teachers. Teacher effectiveness varies little with education level in any grade. Teachers with full credentials are not more effective in any grade than are other teachers.

How does classroom composition affect teacher value added?

Many teachers feel that student performance is based on student background and preparation factors that they are unable to control. The premise is that inner-city teachers serve an at-risk population that will always have lower performing students than their counterparts in more affluent suburbs. This argument has considerable merit for comparing absolute test score levels across schools, but the argument has less merit for comparing improvements in student achievement with value-added models. The value-added approach examines the improvements in student achievement for students assigned to a teacher conditional on their prior achievement scores. The prior scores provide strong evidence of the skills and preparation of each student, so value-added comparisons provide a more meaningful measure of teacher effectiveness than a simple snapshot of how well students perform in one teacher's class.

Table 7 shows how classroom composition affects a teacher's value added scores in reading and math.⁶ We examined the proportion of students with different background characteristics assigned to LAUSD elementary teachers. The average proportions are shown in column 2 of the table. Columns 3 and 4 show how the mix of students assigned to a teacher influence their value-added score.

The results show that most student background factors are unrelated to teacher effectiveness, e.g., students from wealthier families or with better educated parents do not increase teacher value added in either reading or math. These students perform better on achievement tests, but the value added model adjusts for these factors. As a result, most family characteristics do not influence the improvements in test scores that are captured by value added.

Three factors do have a significant effect on value-added scores, but the magnitude of these effects is small. The proportion of gifted students taught by a teacher is positively related to teacher effectiveness, but the effect size is only 0.03 in reading and 0.04 in math. Similarly, the share of black students is negatively related to teacher value added, but the effect size is -0.03 in reading and -0.02 in math. The proportion of Asian/Pacific

⁶ For privacy reasons, the Los Angeles Times did not receive student-level demographic information as part of this study. The results in Table 7 are based on an earlier study by Buddin and Zamarro (2009) that uses similar research methods on data that did include student-level demographic information. These patterns should persist in the current study.

Islander students has no effect on reading scores but a 0.02 effect size in math. Finally, the proportion of ELLs increases teacher value added with an effect size of about -0.01.

These small effect sizes suggest that the value added measure is doing a good job of controlling for the mix of students assigned to individual teachers. While class composition varies considerably across LAUSD, the proportions of students with different demographic and socioeconomic factors have little effect on value added rankings of teacher effectiveness.

Table 7. Value Added and Background of Students Assigned to a Teacher

	Average Proportion	ELA	Math
Female	0.4994	0.0025 (0.0017)	0.0001 (0.0026)
Free/reduced lunch eligibility	0.7733	0.0022 (0.0036)	0.0013 (0.0051)
Gifted	0.1070	0.0301* (0.0031)	0.0390* (0.0050)
Special Education	0.0669	-0.0026 (0.0019)	0.0009 (0.0030)
English Language Learner	0.4671	-0.0092* (0.0033)	-0.0140* (0.0047)
Black	0.1044	-0.0278* (0.0039)	-0.0207* (0.0061)
Hispanic	0.7514	-0.0110 (0.0062)	0.0082 (0.0096)
Asian/Pacific Islander	0.0586	-0.0009 (0.0035)	0.0200* (0.0053)
Parent high school grad	0.2109	-0.0021 (0.0030)	0.0020 (0.0044)
Parent some college	0.1282	-0.0033 (0.0027)	0.0003 (0.0042)
Parent college graduate	0.0991	0.0049 (0.0040)	0.0031 (0.0062)
Parent graduate school	0.0453	0.0056 (0.0037)	0.0030 (0.0047)
Parent education unknown	0.2568	-0.0030 (0.0035)	-0.0037 (0.0053)
Constant		0.0028 (0.0022)	0.0029 (0.0035)
R-squared		0.0767	0.0467
N		9784	9784

Notes: The dependent variables are ELA and math teacher effects from stage 1, adjusted for measurement error. * indicates $p < 0.05$. Robust standard errors are in parentheses. The omitted categories are the proportion of a teacher's students that are male, not eligible for free-reduced lunch, not gifted, not in special education, not ELL, White non-Hispanics, and from a family where neither parent completed high school.

Finally, did students in low-API schools have more ineffective teachers than did students in high-API schools? A one standard deviation change in API was associated with a 0.04 gain in teacher effectiveness in both reading and math. Many teachers in low-API schools are more effective than teachers in high-API schools--about a third of teachers in the lowest API quartile are more effective in reading and math than the typical teacher in the top API quartile. Teachers are slightly more effective in high- than in low-API schools, but the gap is small, and the variance across schools is large.

School Effectiveness

The results in Table 8 show small differences in student achievement from school to school after controlling for lagged achievement and student characteristics. The estimated standard deviation of the school effects is only about 0.06 in ELA and 0.08 in math. After Bayesian adjustment for measurement error, these school effects are about 0.06 in ELA and 0.08 in math.

The estimated school effects are quite small. As discussed above, school effects of 0.06 and 0.08 are small relative to the metrics of annual achievement growth, student achievement gaps between groups, and the effect size of recent school reform programs (Hill et al., 2008). The teacher effects in ELA and math are more than three times as large as the corresponding school effects. The teacher and school results indicate that teacher effectiveness varies much more from classroom to classroom within schools than it does across schools. Effective teachers are not concentrated in a few schools, rather they are spread across the district in low- and high-API schools.

Value-added school effectiveness is positively related to API, but a one standard deviation in school API is only associated with a 0.03 increase in school effects. About a fourth of low-API schools have above average school value added relative to other elementary schools in the district. Similarly, about a fourth of the highest-quartile API schools have below average school effectiveness. The overall message is that many schools with low achievement levels are producing strong achievement gains and many schools with high achievement levels are producing weak achievement gains for their students.

**Table 8: ELA & Math Achievement Regressions
for School Effectiveness**

	ELA	Math
Lagged ELA	0.8872* (0.0010)	
Grade 4	0.0022 (0.0015)	0.0013 (0.0017)
Grade 5	0.0077* (0.0015)	0.0056* (0.0017)
Title I	-0.0400* (0.0034)	-0.0635* (0.0039)
Female	0.0254* (0.0012)	0.0226* (0.0014)
English Language Learner	-0.0325* (0.0015)	-0.0168* (0.0017)
Joined after Kindergarten	0.0238* (0.0014)	0.0176* (0.0016)
Test Year 2005	0.0185* (0.0020)	0.0231* (0.0023)
Test Year 2006	0.0071* (0.0020)	0.0059* (0.0023)
Test Year 2007	0.0071* (0.0021)	0.0091* (0.0024)
Test Year 2008	0.0050* (0.0021)	0.0070* (0.0024)
Test Year 2009	-0.0051* (0.0021)	0.0013 (0.0024)
Lagged Math		0.8847* (0.0011)
Constant	0.0370* (0.0035)	0.0478* (0.0040)
School Effects (σ_τ)	0.0587	0.0842
R-squared	0.6867	0.5970
Student Years	836310	836310
Schools	472	472

Notes: The dependent variables are student-level test scores in ELA and math. * indicates $p < 0.05$.

Robust standard errors are in parentheses. The omitted reference categories are grade 3, not in a Title I school, male, not an ELL, joined LAUSD in kindergarten, and test year 2004. The F test that all $\tau_j = 0$ is 14.70 in ELA and 24.04 in math.

5. CONCLUSIONS AND IMPLICATIONS

The conventional wisdom on what qualifications improve teaching is inconsistent with the empirical results reported here and in several recent studies (Rivkin et al. 2005; Harris and Sass, 2006; Koedel and Betts 2007; Aaronson et al., 2008; and Jacob and Lefgren, 2008). Value-added studies with longitudinal student-level achievement data show that many “important” teacher qualifications have little effect on student outcomes. More experienced or better educated teachers are no more effective in the classroom than inexperienced teachers with only undergraduate diplomas.

New research should focus on measuring teacher skills and preparation that predict subsequent teacher performance in the classroom. The current rules on teacher credentialing and licensure keep many teaching candidates from obtaining certification without much evidence that those candidates would be ineffective in the classroom. Policymakers should carefully consider whether different credentialing practices could improve the quality of the teaching workforce without having severe consequences for teacher supply (Angrist and Guryan, 2003).

The weak effects of measured teacher qualifications 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 improving teacher experience, educational attainment, or licensure scores. A simple reshuffling of teachers is unlikely to produce substantial achievement improvement in low-performing schools. Cash bonuses for these qualifications in low-performing schools will improve the distribution of teacher qualifications across schools without doing much to improve the achievement gap.

Districts could consider developing policies that place importance on output measures of teacher performance. Current policies emphasize teacher qualifications that are inputs to student learning. These inputs are costly to produce and sustain in terms of hiring and salary costs, but they have little consequence on student achievement outcomes. A better approach would be to incorporate value-added measures of teacher effectiveness into teacher assessments. Teachers and administrators should have access to value-added measures of teaching effectiveness. These measures would provide useful feedback for teachers on their performance and for administrators in comparing teacher effectiveness.

Merit pay systems would 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.

Finally, we should remember that the context of teacher assessment and compensation systems affects the relative effectiveness of different types of teachers. We find that teachers with better nominal teaching tools (e.g., experience, education, licensure scores) perform no better than teachers with weaker qualifications, but the current system

provides little reward for better classroom performance. Perhaps teachers with extra teachings skills have too little incentive to fully utilize those skills in a compensation system that rewards their measured inputs and ignores their outputs. By realigning the incentive system and rewarding student achievement gains, we might find a different ordering of teacher effectiveness and improved overall levels of student learning.

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.
- Angrist, J., Guryan, J., 2003. Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements, NBER working paper 9545.
- Borjas, G., 1987. Self-selection and the earnings of immigrants, *American Economic Review*, 77(4), 531-553.
- Buddin, R., McCaffrey, D., Kirby, S., Xia, N., 2007. Merit Pay for Florida Teachers: Design and Implementation Issues, Working paper, RAND WR-508-FEA.
- Buddin, R., Zamarro, G., 2009. Teacher Qualifications and Student Achievement in Urban Elementary Schools, *Journal of Urban Economics*, 66, 103-115.
- Clotfelter, C., Ladd, H., Vigdor, J., 2007. Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects, *Economics of Education Review*, 26, 673-682.
- Gordon, R., Kane, T., Staiger, D. (2006). Identifying Effective Teachers Using Performance on the Job, Brookings Institution Discussion Paper 2006-1.
- Harris, D., Sass, T., 2006. The effects of teacher training on teacher value-added, Working paper, Florida State University.
- Hill, C., Bloom, H., Black, A., Lipsey, M. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research, *Child Development Perspectives*, Volume 2, Number 3, Pages 172–177.
- 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.
- Kane, T., Staiger, D. (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. NBER Working Paper 14607.
- Koedel, C., Betts, J., 2007. Re-examining the role of teacher quality in the educational production function, Working paper, University of California, San Diego.
- Le, V., Buddin, R., 2005., Examining the Validity Evidence for California Teacher Licensure Exams, Working paper, RAND WR-334-EDU.

McCaffrey, D., Sass, T., Lockwood, J., and Mihaly, K. (2009). The Inter-Temporal Variability of Teacher Effects Estimates, *Education Finance and Policy*, Fall 2009, Vol. 4, No. 4: 572–606.

Rivkin, S., Hanushek, E., Kain, J., 2005. Teachers, schools, and academic achievement, *Econometrica*, 73(2), 417-458.

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